Artificial Neural Network Models for Alternative Investments

Paul Merlin

UNIVERSITE DE PARIS I PANTHEON-SORBONNE
U.F.R. DE SCIENCES ECONOMIQUES

Année 2009

Numéro attribué par la bibliothèque

2009PA10020

Thèse

pour obtenir le grade de

Docteur en Sciences Economiques de l'Université de Paris I

« *DES TECHNIQUES NEURONALES DANS L'ALTERNATIF* »

Soutenance publique

le lundi 22 juin 2009 à 14h00

à la Maison des Sciences Economiques

par :

Paul Merlin

(Université Paris-1, CES/CNRS)

Sous la direction de Monsieur le Professeur Thierry CHAUVEAU

(Université de Paris I - CES/CNRS)

Jury de thèse

Directeur de recherche (Paris-1) :	Pr. Thierry Chauveau	(Université de Paris-1)
Président (Paris-1) :	Pr. François Gardes	(Université de Paris-1)
Rapporteur (externe) :	Pr. Sanvi Avouyi-Dovi	(Banque de France)
Rapporteur (externe) :	Pr. Jean-Luc Prigent	(Université de Cergy-Pontoise)
Suffragant (externe) :	Pr. Michel Verleysen	(Université de Louvain)
Suffragant (externe) :	Dr. Robert Kosowski	(*Imperial College* de Londres)
Suffragant (externe) :	Dr. Thierry Roncalli	(Université d'Evry et SGAM)
Suffragant (Paris-1) :	Dr. Bertrand Maillet	(Université de Paris-1)

Je dédie cette thèse, à mes grand-mères, Étiennette et Michelle.

Remerciements

Je souhaiterais, tout d'abord, vivement remercier mon directeur de thèse Monsieur le Professeur Thierry Chauveau. Je lui suis reconnaissant de la confiance dont il a fait preuve en acceptant d'encadrer ce travail de thèse. Ses conseils, ses remarques ainsi que la liberté qu'il a su m'accorder lors du choix des axes de recherche poursuivis ont été particulièrement appréciés.

Je voudrais également exprimer toute ma gratitude à Bertrand Maillet, co-directeur de cette thèse, pour son aide et ses encouragements sans cesse renouvelés. Sa rigueur, sa ténacité et son dynamisme m'ont accompagné toutes ces années.

Je veux remercier vivement Messieurs les Professeurs Sanvi Avouyi-Dovi (Banque de France) et Pr. Jean-Luc Prigent (Université de Cergy-Pontoise) qui ont accepté de me faire l'honneur de rapporter sur cette thèse ; ainsi que Messieurs les Professeurs François Gardes (Université de Paris-1), Michel Verleysen (Université de Louvain) et Messieurs Robert Kosowski (*Imperial College* de Londres) et Thierry Roncalli (Université d'Evry et SGAM) pour avoir, en dépit de leur temps compté, accepté de participer à ce jury.

Ma reconnaissance va ensuite au Professeur Paul Dupuis et toute son équipe de recherche, qui m'ont chaleureusement accueillis, durant l'été 2005, dans leur laboratoire de mathématiques appliquées de l'Université de Brown (USA). Ils m'ont initié aux techniques d'optimisation multi-critères et ils m'ont permis d'avancer de façon significative dans le chapitre 2 intitulé *"Efficient Frontier for Robust Higher-order Moment Portfolio Selection"*.

Je souhaite remercier Amaury Lendasse pour sa chaleureuse invitation à séjourner, au printemps 2006, à l'Université technologique d'Helsinki (Finlande), le berceau des cartes auto-organisatrices de Kohonen. Ce séjour a constitué une formidable opportunité pour me familiariser à cetaines techniques neuronales. Il m'a également permis de débuter une collaboration fructueuse qui, je l'espère, durera.
Mes remerciements vont également au Professeur Stephen Brown pour m'avoir convié, durant l'automne 2006, aux séminaires de finance de l'Université Stern de New York (USA). Son aide et ses remarques concernant le chapitre 4, intitulé "Carte de style et facteurs de risques", ont été particulièrement enrichissantes.

Je pense également à Thierry Michel, Emmanuel Jurczenko, Christophe Boucher et Jean-Philippe Médecin, pour m'avoir soutenu et éclairé tout au long de ce travail de recherche. Leurs remarques, suggestions et sagacités ont été déterminantes dans l'accomplissement de ce travail.
Je tiens ensuite à remercier tout particulièrement Patrick Kouontchou, Benjamin Hamidi, Guillaume Bagnarosa, Grégory Jannin pour leur assistance ainsi que l'atmosphère studieuse et décontractée qu'ils ont contribué à instaurer.
Je voudrais aussi remercier vivement toute l'équipe A.A.Advisors (ABN AMRO), et plus particulièrement Monsieur Stéphane Corsaletti, pour m'avoir permis de travailler au sein de ses services de recherche quantitative et de gestion des risques depuis le début de cette thèse. Ces expériences m'ont donné l'opportunité de mettre en pratique mes connaissances de la gestion alternative et d'acquérir les compétences d'un *Risk Manager* certifié.
J'adresse également des sincères remerciements aux participants des conférences et séminaires auxquels j'ai participé : Conférence internationale de l'AFFI (Paris, Juin 2005 ; Bordeaux, Juin 2007), *Workshop On Self-organizing Maps* (WSOM - Paris, Septembre 2005 et St. Augustine - USA, Juin 2009), *International Conference On Artificial Neural Networks* (ICANN - Varsovie, Septembre 2005), Journées de Micro-économie Appliquée (JMA06 - Nantes, Juin 2006), *European Symposium on Artificial Neural Networks* (ESANN - Bruges, Avril 2007 et Avril 2009), Journée d'Econométrie Financière Avancée (Paris, Décembre

2005, Décembre 2006, Décembre 2007 et Décembre 2008), Conférence de l'Association Franaise de Science Economique (AFSE09 - Paris, Septembre 2009). Qu'ils trouvent ici une trace des remarques pertinentes formulées lors de mes présentations.

Je précise ici que les travaux réalisés en vu de la rédaction du chapitre 1, intitulé "*Hedge Funds Time Series Completion and* Scenarii *Generation for Robust Asset Allocation and Risk Measurement*", m'ont permis de réaliser une série d'articles dont certains ont été rédigés en collaboration avec Amaury Lendasse, Bertrand Maillet et Antti Sorjamaa. De même, les résultats obtenus dans le chapitre 2, intitulé "*Efficient Frontier for Robust Higher-order Moment Portfolio Selection*", sont présentés dans un article co-écrit avec Emmanuel Jurczenko et Bertrand Maillet. Mes co-auteurs m'ont largement fait bénéficier, à cette occasion, de leur connaisance approfondie dans le domaine particulier des modèles multi-moments. Le chapitres 3, intitulé "*Outliers Detection, Correction of Financial Time-series Anomalies and Distributional Timing for Robust Efficient Higher-order Moment Asset Allocations*" ainsi que le chapitre 4, "Carte de style et facteurs de risques", devraient également aboutir, dans un avenir proche, à la soumission d'articles co-écrits avec Christophe Boucher et Bertrand Maillet. Je pense ici à mes co-auteurs avec qui j'ai eu le plaisir d'échanger et d'apprendre ces dernières années.

Mes remerciements vont enfin à mes proches : ma soeur et mes frères, Sabine, Bruno et Clément. A mes amis, Charlie Cussac, Clara et Jérôme Sitruk, Jérémie Anquez, Julie Perier, Mihai Gavriloiu et bien d'autres encore. Je pense tout particulièrement à Nolwenn Guellec qui a su me soutenir lors de certains moments difficiles. Qu'ils sachent tous que leur soutien a également été déterminant pour la réalisation de ce travail.

Je ne saurais conclure sans avoir une pensée à tous ceux ou celles qui ont contribué à améliorer ce travail. Ce travail a bénéficié, dans sa version finale, des remarques et suggestions de plusieurs lecteurs. Néanmoins, les éventuelles erreurs, omissions ou coquilles ne pourront pas leurs être incombées. Elles resteront donc entièrement à la charge de l'auteur.

Table des matières

Remerciements **iii**

Table des figures **xii**

Liste des tableaux **xv**

Introduction Générale **1**

Résumé du chapitre sur "*Hedge Fund Time-series Completion and* Scenarios *Generation for Robust Asset Allocation and Risk Measurement*" **10**

Résumé du chapitre sur "*Efficient Frontier for Robust Higher-moment Portfolio Selection*" **21**

Résumé du chapitre sur "*Outliers Detection, Correction of Financial Time-series Anomalies and Distributional Timing for Robust Efficient Higher-order Moment Asset Allocations*" **32**

Résumé du chapitre sur "Carte de style et facteurs de risques" **43**

1 Hedge Fund Time-series Completion and *Scenarii* Generation for Robust Asset Allocation and Risk Measurement **55**

1.1 Introduction 55

1.2 Imputation Methods 58

1.2.1 Self-Organizing Map Algorithms 59

1.2.2 Empirical Orthogonal Functions 63

1.2.3 Robust Flexible Completion 64
1.3 *Scenarii* Generation 66
1.4 Empirical Illustrations 68
1.4.1 Impact of Estimation Errors on Asset Allocation 70
1.4.2 Impact of Estimation Errors on Risks Measures 72
1.5 Conclusion 75
1.6 Appendix 75
1.6.1 HFR^{TM} Main Classification - Strategies Definitions 75
1.6.2 Expected Maximization Methods 77
1.6.3 VaR Methodologies 79
1.6.4 Value-at-Risk Estimation 81

2 Efficient Frontier for Robust Higher-order Moment Portfolio Selection 83
2.1 Introduction 83
2.2 Robust Higher-order Moments for Portfolio Selection 88
2.2.1 Population L-comoments 89
2.2.2 Sample Portfolio Return L-moments 93
2.3 Portfolio Selection with Higher-order Moments 96
2.3.1 Higher-order C-comoments of Portfolio Returns 99
2.3.2 Higher-order L-comoments of Portfolio Returns 104
2.3.3 Higher-order L-moments and the Efficient Frontier Definition 107
2.3.4 The Shortage Function and the Robust Efficient Frontier 109
2.4 Data and Empirical Results 114
2.5 Conclusion 131
2.6 Appendix 134
2.6.1 Appendix 1 134
2.6.2 Appendix 2 137
2.6.3 Appendix 3 140

3 Outliers Detection, Correction of Financial Time-series Anomalies and Distributional Timing for Robust Efficient Higher-order Moment Asset Allocations 145

3.1 Introduction . . . 145
3.2 Detecting and Correcting the Outliers . . . 149
3.2.1 From GARCH to ANN-GARCH Modelling . . . 149
3.2.2 GARCH-related Outlier Detection Models . . . 153
3.2.3 From Detection to Correction of Outliers . . . 156
3.3 Portfolio Selection with Higher-order Moments . . . 158
3.3.1 Higher-order C-comoments of Portfolio Returns . . . 158
3.3.2 Higher-order L-moments of Portfolio Returns . . . 160
3.3.3 The Shortage Function and the Efficient Frontiers . . . 162
3.3.4 Distributional Timing and Higher-order Moments . . . 164
3.4 Data and Empirical Results . . . 167
3.4.1 About the Detection of Market Data Outliers on the French Market 168
3.4.2 About Financial Features of some Notorious Efficient Portfolios . . 172
3.5 Conclusion . . . 180
3.6 Appendices . . . 181
3.6.1 Appendix 1: About Robust Estimation of GARCH(1,1) Models . . 181
3.6.2 Appendix 2: Gaussianity Tests . . . 183
3.6.3 Appendix 3: Utility-based Performance Measures . . . 185

4 Carte de style et facteurs de risque 192

4.1 Introduction . . . 192
4.2 L'analyse de style . . . 195
4.2.1 Les modèles d'évaluation . . . 195
4.2.2 Les modèles d'analyse de style . . . 197
4.3 Les facteurs de style . . . 202
4.3.1 Cartes auto-organisées de Kohonen . . . 202
4.3.2 Cartes robustes . . . 204

4.3.3 Analyse de style par cartes robustes 205
4.4 Détermination des styles par cartes robustes sur le marché des fonds d'actions américaines 208
4.4.1 Création et validation de la carte 209
4.4.2 Interprétation de la carte des fonds en terme de styles : approches classique et conditionnelle 212
4.4.3 Recherche des biais dans l'auto-déclaration des styles des gérants . . 217
4.4.4 Stabilité de la carte des styles américains et dynamique des gérants 218
4.5 Conclusion 222
4.6 Annexe 223
4.6.1 Annexe 1 : Construction des indices de style 223
4.6.2 Annexe 2 : Etude préliminaire, calibration de la carte et choix des inputs 225

Conclusion Générale **235**

Références bibliographiques **243**

Annexe : Articles complémentaires **279**

Annexe 1 : *"Completing Hedge Fund Missing Net Asset Values using Kohonen Maps and Constrained Randomization"* 279
Annexe 2 : *"Robust SOM for Realistic Data Completion"* 285
Annexe 3 : *"SOM+EOF for Finding Missing Values"* 293
Annexe 4 : *"Hedge Funds Portfolio Selection with Higher-order Moments: A Non-parametric Mean-Variance-Skewness-*Kurtosis *Efficient Frontier"* . . . 299
Annexe 5 : *"A Robust Hybrid DHMM-MLP Modelling of Financial Crises measured by the WhIMS"* 314
Annexe 6 : *"X-SOM and L-SOM: A Nested Approach for Missing Value Imputation"* 321

Annexe 7 : *"A Non-linear Approach for Completing Missing Values in Temporal Databases"* 325

Table des figures

1.1 Illustration of the Kohonen Iterative Process. 60
1.2 Step 1, Bootstrap Process for Building the Table **P** of the Individual's. . . 61
1.3 Step 2, RMap Selection. 62
1.4 Representation of the RFC 66
1.5 Error Densities of the First Four Moments. 69
1.6 Efficient Frontier after Completion. 71
1.7 Density Estimation of HFR Global Index. 73

2.1 Comparison of Classical Moments *versus* L-moments. 95
2.2 First Four L-moment Constrained Efficient Frontier in the L1-L2 Plane. . . 117
2.3 First Four L-moment Constrained Efficient Frontier in the L2-L3 Plane. . . 118
2.4 First Four L-moment Constrained Efficient Frontier in the L2-L4 Plane. . . 119
2.5 First Four L-moment Constrained Efficient Frontier in the L1-L4 Plane. . . 120
2.6 First Four L-moment Constrained Efficient Frontier in the L1-L3 Plane. . . 120
2.7 First Four L-moment Constrained Efficient Frontier in the L3-L4 Plane. . . 121
2.8 First Four L-moment Constrained Efficient Frontier in the L1-L2-L3 Space. 123
2.9 First Four L-moment Constrained Efficient Frontier in the L1-L2-L3 Space. 124
2.10 Clusters of the Four L-moment Optimal Portfolios according to their L-moments in the L1-L2 Plane. 126
2.11 Optimal Portfolios for various Four-moment Dependent Utility Functions in the L1-L2 Plane. 128
2.12 Efficient Frontiers for CRRA Utility Functions with Intensified Higher-order Moments. 130

2.13 Set of Portfolios with Second L-moments almost Equivalent to the one of the Minimum L-moment2 Portfolio. 132
2.14 Clusters of the Four L-moment Optimal Portfolios according to their L-moment Rankings in the L1-L2 Plane. 144

3.1 Simplified Representation of a Typical MultiLayer Perceptron. 152
3.2 French Equity Market Evolution 168
3.3 Comparison of ANN-GARCH and GARCH detected Outliers 170
3.4 Time-evolutions of some Notorious Efficient Portfolios 174
3.5 Subjective *versus* Objective Probability Functions of the Returns on the Global Minimum Volatility Portfolio 191

4.1 Représentation du processus itératif d'apprentissage de cartes auto-organisées. 203
4.2 Représentation de l'algorithme de cartes robustes. 205
4.3 Evolution des performances des fonds du marché (absolues à gauche, relatives à droite. 208
4.4 Evolution des performances des indices de style du marché. 209
4.5 Performances relatives représentatives des seize classes (pour chaque cluster à gauche, superposées à droite). 210
4.6 Carte des distances séparant les individus représentatifs des classes. 211
4.7 Repérage des indices représentatifs et des facteurs sur la carte. 213
4.8 Evolution des corrélations des indices de style *Large* et *Small*, et *Growth* et *Value*. 215
4.9 Affection de biais au moyen de classifications conditionnelles. 216
4.10 Superposition des auto-déclarations des styles et la classification en rendements relatifs obtenue *via* une carte robuste. 218
4.11 Représentation du parcours du fonds Russell Diversified Equity Fund sur la carte. 219
4.12 Comparaison des parcours de deux fonds sensés appliquer un même style de gestion. 220

4.13 Histogramme des scores de dynamique des fonds. 222
4.14 Représentation des différentes fonctions de voisinage. 229
4.15 Evolutions des valeurs liquidatives des individus représentatifs des seize classes suivant le type de classification utilisée. 232

Liste des tableaux

1.1 Mean Errors on the Statistics of rebuilt Time Series. 69
1.2 Root Mean Square Error of Value-at-Risk obtained after Completion. . . . 73
1.3 Absolute Error of Monte Carlo Value-at-Risk obtained for the Maximum Sharpe Portfolio. 74
1.4 Absolute Error of Monte Carlo Value-at-Risk obtained for the equaly-weigthed portfolio. 74
1.5 Value-at-Risk 99%. 81
1.6 Value-at-Risk 95%. 82

2.1 P-statistics of Normality Goodness-of-Fit Tests of the various Optimal Portfolio Return Series (frequencies of rejections at 1%, 5% and 10% confidence levels) 125
2.2 Statistical Properties of Minimum L-moment$_2$ Portfolio versus the Local Maximum Mean, Local Maximum L-moment$_3$ and Local Minimum L-moment$_4$ Portfolios 133

3.1 Descriptive Statistics 169
3.2 Descriptive Statistics of the Outliers 171
3.3 Largest Outliers on the Database 171
3.4 Local Maximum Mean Portfolios 175
3.5 Performance Measures for Local Maximum Mean Portfolios 175
3.6 Global Minimum Volatility Portfolios 176
3.7 Performance Measures for Global Minimum Volatility Portfolios 177
3.8 Local Maximum Skewness Portfolios 178

3.9 Performance Measures for Local Maximum Skewness Portfolios 178
3.10 Local Minimum *Kurtosis* Portfolios 179
3.11 Performance Measures for Local Minimum *Kurtosis* Portfolios 179
3.12 First Four Moments of Stocks of the Sample 184
3.13 Statistical Properties of Stock Returns 185
3.14 Heteroskedasticity Tests of Stock Returns 186

4.1 Caractéristiques linéaires de la classification. 212
4.2 Matrice de corrélation des indices de style. 214
4.3 Taux d'erreur des styles auto-déclarés. 218
4.4 Dynamique des fonds, changement de classe des fonds. 221
4.5 Score dynamique des fonds. 221
4.6 Mesures d'appréciation des cartes en fonction du type de données choisi exprimées en pourcentages 233
4.7 Mesure de la qualité des classifications obtenues en rendements relatifs (exprimées en pourcentages) 234

Introduction Générale

Depuis le début des années 1990, l'industrie de la gestion alternative a connu un développement considérable. La croissance spectaculaire des actifs sous gestion – de quelques milliards de dollars en 1995, à plus de 2 600 milliards de dollars en 2007 (*Cf. HedgeFund Intelligence*, 2008) – en fait désormais un acteur incontournable de la sphère financière. L'influence de ces fonds alternatifs (en anglais, *hedge funds*), présents sur les principaux marchés financiers, va bien au-delà de leurs poids financiers. Leur activisme dans les conseils d'administration des entreprises et l'utilisation d'effet de levier décuplent l'impact des *hedge funds* sur la sphère financière classique.

Définir précisément la gestion alternative n'est pourtant pas chose aisée (*Cf.* Hilderbrand, 2007). L'Autorité des Marchés Financiers (AMF, l'organisme de régulation des marchés français) définit la gestion alternative comme une gestion qui a pour objectif la performance absolue en contrôlant son exposition aux facteurs traditionnels de risque. Selon la *Securities and Exchange Commission* (*SEC*, l'organisme de régulation des marchés américains), un *hedge fund* est un véhicule privé et non référencé de fonds commun de placement qui utilise des techniques sophistiquées de couverture et d'arbitrage. L'univers de la gestion alternative regroupe un ensemble de fonds (pas ou peu contrôlés par les autorités de tutelle), qui peuvent de fait adopter des stratégies d'investissement très diverses. Les fonds alternatifs ont en outre la possibilité d'investir dans de nombreuses classes d'actifs et d'utiliser un vaste ensemble d'outils financiers. Afin d'éviter certaines contraintes fiscales et réglementaires, la structuration des *hedge funds* est généralement complexe.

Un manque de transparence est ainsi souvent reproché aux acteurs de la gestion alternative. En effet, les informations relatives aux stratégies misent en œuvre par les gérants sont

délibérément protégées et confidentielles, ce qui entretient une opacité générale de cette industrie. Les gestionnaires justifient cette rétention d'information par des arguments d'exclusivité et de viabilité de leurs stratégies. Les gérants cherchent à exploiter d'éventuelles opportunités d'arbitrage ; la divulgation de leurs stratégies entraînerait une correction de ces inefficiences, et donc, la disparition de leurs « martingales ».

Le terme « *hedge* » (couverture), à ce propos, peut induire en erreur. Le premier *hedge fund* aurait été fondé par Alfred Winslow en 1949 (*Cf.* Lhabitant, 2002). Ce fonds autorisait une couverture du risque de marché en adoptant une stratégie d'achats et de ventes simultanés d'actions (en anglais *Long/Short*). Ainsi, les *hedge funds* ne se couvrent pas, bien sûr, contre tous les risques financiers (*Cf.* Asness *et alii*, 2001 ; Fromont, 2006), mais ils tentent néanmoins d'en maîtriser une partie, en ne s'exposant qu'à certains facteurs de risque choisis.

L'engouement des investisseurs pour ce type de produits s'explique en partie par la diversification qu'ils apportent (*Cf.* Amin et Kat, 2003-a). Les perspectives de gains importants pour un risque visiblement modéré en font de plus un type de produit attrayant (*Cf.* Amin et Kat, 2003-b). Nombre d'études (*Cf.* Agarwal et Naik, 2004 par exemple) tendent à montrer que l'allocation d'une poche alternative au sein d'un portefeuille d'actifs ou de fonds plus classiques permet d'espérer un portefeuille de risque global moindre pour un rendement équivalent.

Par ailleurs, les méthodes d'analyses financières « classiques » ne s'accordent pas toujours avec les spécificités des fonds alternatifs. La plupart des modèles financiers supposent des dépendances linéaires entre les rentabilités des fonds et les facteurs de risque. Ces modèles se fondent souvent sur l'hypothèse de normalité de la distribution des rendements des actifs. Or, les particularités structurelles des fonds alternatifs (utilisation de levier, de produits dérivés, investissement dans des titres peu liquides, valorisation de faible fréquence, existence de délais de sortie...) rendent difficiles l'évaluation de la performance et des risques réels (*Cf.* Fung et Hsieh, 2000). La volatilité, par exemple, ne reflète que (très) partiellement le risque d'un fonds alternatif. L'utilisation d'actifs non liquides

et de produits dérivés, ainsi que l'application de stratégies dynamiques d'exposition aux risques (*Cf.* Gibson et Gyger, 2007), ont un impact important et significatif sur la distribution des rendements. Les expositions non-linéaires des fonds à certains facteurs de risque induisent des rendements par nature non-Gaussiens (*Cf.* Agarwal et Naik, 2000 ; Amin et Kat, 2003-b). Le cadre d'analyse classique du choix de portefeuille (moyenne-variance) proposé par Markowitz (1952) n'est en conséquence pas adapté au traitement de tels actifs. Les modèles de risques et d'allocation d'actifs qui découlent de ce cadre d'analyse ne sont alors plus réalistes. Cette thèse ambitionne de proposer un nouveau cadre d'analyse moins restrictif et compatible avec les caractéristiques des fonds alternatifs. Nous proposons en particulier d'élargir aux quatre premiers moments l'analyse traditionnelle moyenne-variance. La considération des moments d'ordre supérieur et plus généralement des dépendances non-linéaires des actifs complexifient néanmoins grandement les modélisations financières. Il n'est alors pas toujours possible d'obtenir des formules explicites « fermées » en considérant des approches mathématiques traditionnelles. Nous choisissons ainsi dans la suite de la thèse d'employer des « procédés de calcul naturel » (en anglais, *natural computation*) pour réaliser certaines approximations et simplifier les problèmes complexes rencontrés.

Les techniques de calcul naturel regroupent trois grandes classes de méthodes ; leur caractéristique commune est d'être en lien direct avec des manifestations de phénomènes naturels (biologiques, physiques, écologiques… *; Cf.* de Castro, 2007). La première classe s'inspire de phénomènes naturels pour réinventer des méthodes de résolution de problèmes. La deuxième se fonde sur une utilisation intensive de l'informatique pour reproduire ou synthétiser ces phénomènes. La troisième utilise des éléments naturels (molécules ou neurones biologiques par exemple) pour développer de nouvelles techniques numériques. Par exemple, les algorithmes génétiques utilisés pour rechercher des solutions optimales sont directement issus des théories de l'évolution de Darwin. Ces techniques inspirées de la nature font désormais l'objet d'intenses développements. Elles constituent un champ actif de recherche en mathématiques appliquées, dont les résultats sont régulièrement pu-

bliés dans des revues scientifiques spécialisées (*Natural Computing*, *Neurocomputing*, *Neural Networks*, *IEEE Transactions on Neural Networks*, *IEEE Transactions on Knowledge and Data Engineering*, *Artificial Intelligence in Medicine*, *Neural Processing Letters*...) et lors de conférences internationales (*Neural Information Processing Systems*, *International Conference on Artificial Neural Networks*, *European Symposium on Artificial Neural Networks*, *International Work-Conference on Artificial Neural Networks*, *Workshop on Self-Organizing Maps*, *Interdisciplinary Center for Neural Computation*).

Le cerveau humain reste une des sources prépondérantes d'inspiration des procédés de calcul naturel. McCulloch et Pitts (1943) modélisent les premiers neurones biologiques à partir d'automates binaires. Ils montrent notamment qu'interconnectés en réseau, ces simples neurones peuvent reproduire les résultats de fonctions relativement complexes. Une deuxième génération de neurones, intégrant notamment une fonction d'activation non-linéaire a permis de reproduire partiellement les facultés d'analyse et d'apprentissage de leurs homologues naturels (*Cf.* Rosenblatt, 1958). L'invention de la rétropropagation du gradient (apprentissage non-supervisé ; *Cf.* Rumelhart *et alii*, 1986) et la démonstration du théorème d'approximation universelle (un type spécifique de réseaux de neurones peut en théorie approximer n'importe quelle fonction ; *Cf.* Hornik *et alii*, 1989) ont permis une plus importante diffusion de ces méthodes au sein de la communauté scientifique.

Combinés en réseau, ces modèles connexionnistes offrent des possibilités intéressantes par rapport aux autres modèles linéaires ou statistiques classiques. Les problèmes de projection et de réduction de dimension, souvent traités au moyen d'Analyse en Composantes Principales (*Cf.* Jolliffe, 1986), peuvent être résolus au moyen de réseaux de neurones spécifiques (*Cf.* Lee et Verleysen, 2005). Les modélisations de séries temporelles peuvent également être réalisées efficacement à l'aide de techniques connexionnistes (*Cf.* Lendasse *et alii*, 2000). Ils ont ainsi été appliqués à de très nombreux domaines : optimisation et prévision (trafic de réseau routier, consommation d'électricité), reconnaissance (vocale ou optique), approximation (de fonctions inconnues ou complexes), ou compression (d'image ou de données).

Dans le domaine financier, dès 1990, Hawley *et alii* appliquent ces méthodes pour traiter des problèmes aussi divers que la détection des entreprises en difficulté, la gestion de portefeuille, l'évaluation des introductions en bourse et la détection d'opportunités d'arbitrage. Aujourd'hui encore, les vastes possibilités offertes par ces techniques sont communément employées par la recherche académique en finance. Elles sont par exemple utilisées pour prévoir le prix d'actifs (*Cf.* Avouyi-Dovi et Caulet, 1995), pour discriminer différents actifs financiers (*Cf.* Maillet et Rousset, 2003 ; de Bodt *et alii*, 2004), ou encore pour corriger les variations saisonnières des séries temporelles (*Cf.* Ben Omrane et de Bodt, 2007). Pourtant, le caractère non-supervisé de ces modèles (et leur caractéristique dite de « boîte noire ») constitue une critique récurrente. Néanmoins, une utilisation circonspecte et des contrôles *a posteriori* (*Cf.* de Bodt *et alii*, 2002 et 2004) permettent de concevoir raisonnablement des modèles fondés sur ces techniques neuronales. Nous envisageons ainsi d'exploiter les propriétés des réseaux de neurones pour intégrer les spécificités des rendements des fonds alternatifs.

Notons que le cadre d'analyse que nous proposons s'avère également particulièrement utile lorsqu'il s'agit d'étudier les rentabilités des actifs plus classiques. Car si les rendements des fonds alternatifs se caractérisent par leur non-normalité, l'hypothèse de rendements Gaussiens pour les actifs financiers est aussi largement rejetée par la littérature empirique (*Cf.* Fama, 1965 ; Mandelbrot, 1997). Le caractère asymétrique et leptokurtique des rentabilités des actifs financiers – certes moins prononcé que pour les fonds alternatifs – a en effet suscité plusieurs développements et extensions du cadre d'analyse classique (*Cf.* Bollerslev *et alii*, 1992 ; Pagan, 1996 ; Harvey et Shephard, 1996 ; Campbell *et alii*, 1997 ; Cont, 2001 et Gouriéroux et Jasiak, 2001 ; Jondeau et Rockinger, 2003-a).

En premier lieu, les queues de distributions empiriques sont plus épaisses que celles d'une loi Normale. Les distributions empiriques sont également asymétriques ; il est ainsi plus probable d'observer des rendements extrêmes négatifs que positifs. Ensuite, les phénomènes de corrélation sérielle de la volatilité, qui se manifestent par l'apparition de « grappes » de volatilités élevées (en anglais, *volatility clusters*), ainsi que les évolutions des corrélations transversales (les corrélations ont tendance à augmenter durant les périodes de fortes vo-

latilités, *Cf.* Gagliardini et Gouriéroux, 2006) vont également à l'encontre du postulat Gaussien. Si ces faits stylisés sont aujourd'hui bien connus, leurs justifications restent encore partielles. Ainsi, les phénomènes de bulles spéculatives (*Cf.* Blanchard et Watson, 1982 ; Abreu et Brunnermeier, 2003) et l'existence de modes d'incitation optionnelle (*Cf.* Black et Scholes, 1973 ; Brennan, 1993 ; Bris *et alii*, 2007) induisent une asymétrie dans les distributions de rentabilité des actifs. Par ailleurs, les phénomènes de contagion, d'hétéroscédasticité des rentabilités, ainsi que l'hétérogénéité des anticipations des agents (*Cf.* Bollerslev, 1986 ; Hamao *et alii*, 1990 ; Embrechts *et alii*, 1999 ; Hong et Stein, 1999 et 2003 ; Chen *et alii*, 2001 ; Lillo *et alii*, 2009), sont à l'origine du caractère asymétrique et leptokurtique des fonctions de densité des rentabilités des actifs. Enfin, l'existence de risque de crédit et de liquidité peut être à l'origine de certains phénomènes de pertes extrêmes (*Cf.* Kritzman, 1994). Les évènements de marché survenus au cours du quatrième trimestre 2008 en sont une parfaite illustration.

Cette thèse cherche à répondre à un certain nombre de questions auxquels les praticiens et les chercheurs académiques sont souvent confrontés lors de l'application de modèles financiers à des classes d'actifs comprenant des fonds alternatifs. Plus précisément, nous présentons dans un premier chapitre une méthode de prétraitement des bases permettant d'estimer les valeurs manquantes des séries d'actifs. Le second chapitre propose ensuite un modèle de sélection de portefeuilles optimaux intégrant la non-normalité des actifs. Le troisième chapitre traite de la problématique des valeurs aberrantes (détection et correction) présentes dans des bases de données financières. Enfin, un quatrième chapitre discute d'un modèle de caractérisation des styles des fonds alternatifs prenant en compte les dynamiques de gestion.

Dans le premier chapitre, nous cherchons à répondre à la problématique de génération de rendements réalistes. L'objectif est double : nous proposons d'une part une méthode d'estimation des valeurs manquantes (également appelée méthode de complétion) qui peuvent (ne pas) apparaître dans les bases de données financières pour de multiples rai-

sons ; et d'autre part, une méthode de génération de *scenarii* réalistes de marché. En effet, qu'il s'agisse de méthode de complétion ou de simulation, l'hypothèse distributionnelle de rendements normaux (ou symétriques) était jusqu'à présent généralement retenue. L'utilisation de modèles standard pour des classes d'actifs dont les rendements sont jugés non-normaux n'est pourtant pas recommandée, car elle induit des biais d'analyse marqués. Il est ainsi nécessaire de développer des modèles adaptés aux spécificités des fonds alternatifs. Nous proposons de compléter les valeurs manquantes en associant trois approches complémentaires : une version modifiée de l'algorithme de Kohonen, les « Cartes Robustes Auto-Organisées » (en anglais *Robust Self-Organized Maps* ou *RSOM* ; *Cf.* Guinot *et alii*, 2006), une reconstruction factorielle à partir des Fonctions Orthogonales Empiriques (en anglais *Empirical Orthogonal Functions* ou *EOF* ; *Cf.* Hanson et Lawson, 1974), et enfin un algorithme de génération aléatoire contrainte de séries temporelles (en anglais *Constrained Randomization* ; *Cf.* Schreiber, 1998). Nous adoptons ensuite une approche non paramétrique qui se fonde, à nouveau, sur les cartes robustes pour générer des *scenarii*. Les propriétés d'interpolation non-linéaire des cartes nous permettent d'obtenir des *scenarii* respectant les distributions et les dépendances des actifs à simuler. Nous illustrons aussi l'impact de la complétion de bases de fonds alternatifs sur des mesures de risque et des allocations d'actifs et nous les comparons à celles issues de techniques plus classiques.

Le deuxième chapitre présente une généralisation de la frontière efficiente proposée par Markowitz (1952). Motivé par le caractère non-Gaussien des distributions des rentabilités des actifs financiers (*Cf.* Fama, 1965 ; Mandelbrot, 1997) et plus particulièrement des fonds alternatifs, nous étendons le cadre initial espérance-variance d'évaluation des portefeuilles aux moments d'ordre supérieur qui caractérisent les cœfficients d'asymétrie et d'épaisseur de queues de distribution des rendements (en anglais, *skewness* et *kurtosis*). Néanmoins, les moments d'ordre trois et quatre étant particulièrement sensibles aux données aberrantes (*Cf.* Kim et White, 2004), nous proposons d'utiliser des statistiques robustes, « les moments linéaires » (en anglais *L-moments* ; *Cf.* Hosking, 1990), pour quantifier l'asymétrie des distributions et les risques extrêmes associés aux rentabilités de portefeuilles. L'intro-

duction d'une mesure directionnelle de distance (directement liée aux préférences de l'investisseur) autorise ensuite la résolution d'un programme de maximisation multi-objectif (considérant simultanément espérance, variance, asymétrie et *kurtosis*). La considération de l'ensemble des distances, correspondant à des préférences « raisonnables » d'agents rationnels, nous permet d'obtenir l'ensemble des portefeuilles (localement) optimaux. Enfin, l'introduction de fonctions d'utilité générales, intégrant notamment les préférences des agents pour les moments d'ordre supérieur, nous permet d'évaluer l'influence de ces moments sur les choix d'investissement des agents rationnels.

Dans le troisième chapitre, nous proposons un modèle de « débruitage » des séries financières. Initialement développé en traitement du signal, ce procédé consiste à supprimer le bruit (la partie résiduelle qui perturbe un signal, une image, un son…, les séries de rendements dans notre cas). Au moyen d'un réseau de neurones modélisant les volatilités instantanées et transversales d'une base d'actifs, nous détectons et nous corrigeons les valeurs aberrantes. Ce prétraitement des bases semble en effet utile. Comme le montre Rosenberg et Houglet (1974), quelques erreurs dans les données suffisent à biaiser largement les conclusions issues des modélisations financières. Les allocations d'actifs obtenues dans le cadre espérance-variance proposé par Markowitz (1952) sont particulièrement sensibles à ces problèmes d'estimation (*Cf.* parmi d'autres Britten-Jones, 1999). De nombreux auteurs (*Cf.* Michaud, 1998 ; Ledoit et Wolf, 2003) proposent d'ailleurs des solutions robustes pour déterminer les portefeuilles optimaux. Deux types d'approche sont ainsi possibles. La première consiste à effectuer un prétraitement de la base de données ; il va s'agir de détecter et corriger les rendements « anormaux ». La seconde substitue aux moments empiriques traditionnels (sujets aux erreurs d'estimation) des statistiques robustes. L'asymétrie et la *kurtosis* étant particulièrement sensibles à ces données erronées, il nous semblait intéressant d'étudier l'impact de ces données « aberrantes » sur les modèles d'allocation de portefeuilles intégrant ces moments d'ordre supérieur. Nous proposons ainsi de comparer les allocations de portefeuilles obtenues à partir du modèle proposé par Jurczenko *et alii* (2006) - lorsque les données sont brutes ou débruitées - et celui proposé au

chapitre 2 mobilisant des statistiques robustes.

Enfin, dans le quatrième chapitre, nous développons un modèle d'analyse de style adapté aux fonds alternatifs. Généralement supposés constants, les styles adoptés par les gérants sont susceptibles d'évoluer en fonction des conditions de marché. Cette dérive peut être volontaire (choix tactique du gérant) ou simplement subie (*Cf.* Fama et French, 2007) en raison de l'évolution des caractéristiques intrinsèques des actifs détenus, le style retenu par le gérant étant un élément primordial pour l'évaluation de la performance des fonds. La littérature académique s'est ainsi largement intéressée à l'identification du style des fonds et l'évaluation des biais concernant les styles auto-déclarés des gérants. Nous développons un modèle d'analyse de style dynamique fondé sur les cartes auto-organisées de Kohonen (1995). Cette approche permet d'éviter certains écueils des modèles traditionnels de l'analyse de style (*Cf.* Sharpe, 1988), comme l'arbitraire du choix des indices, la colinéarité des indices de style retenus et les hypothèses liées aux méthodes d'estimation du modèle de régression multivariée. De façon préliminaire, nous proposons d'appliquer notre nouvelle méthodologie à un univers de fonds d'actions américaines, pour lesquels un grand nombre d'informations sont publiques (le style auto-déclaré du gérant, les principales positions, la classification attribuée par le fournisseur de données, *etc.*). Une projection dynamique des fonds sur la carte rend alors possible l'étude des styles adoptés par fonds et leur évolution. Nous déterminons ainsi deux types de fonds : ceux qui adoptent une stratégie stable et cohérente sur la période, et ceux, plus opportunistes, qui changent régulièrement de styles. Parmi ces derniers, nous tentons d'identifier les gérants qui anticipent avec succès les futures conditions de marché de ceux les subissant.

Avant de présenter en détail les travaux réalisés pour traiter ces différentes problématiques, nous proposons ci-après un résumé détaillé des quatre parties de la thèse.

Résumé du chapitre sur "*Hedge Fund Time-series Completion and* Scenarii *Generation for Robust Asset Allocation and Risk Measurement*"

Les études empiriques constituent une étape incontournable de la recherche en finance. Elles permettent, entre autres, de valider ou d'appliquer les résultats de modèles théoriques. Disposer d'observations « propres » et en nombre suffisant est un préalable nécessaire à l'application de ces modèles. Aussi, l'existence de données manquantes constitue un problème récurrent pour nombre d'analyses (*Cf.* Malhotra, 1987 ; Schafer, 1997 ; diCesare, 2006). Outre ces problèmes de valeurs manquantes, d'autres considérations empiriques, tel qu'un nombre d'observations trop restreint, limitent le pouvoir explicatif de certains modèles. Le recours à la simulation et aux techniques de générations de *scenarii* s'impose alors. Ce premier chapitre de notre travail de thèse propose dans un premier temps un modèle de reconstruction des valeurs manquantes dans des séries temporelles. Dans un second temps, la généralisation de ce principe nous permet de proposer un modèle de génération de *scenarii* réalistes de marché.

Rubin (1976) propose l'une des premières analyses systématiques des valeurs manquantes. Parmi les diverses méthodes de traitement proposées, on distingue généralement deux grandes catégories. La première consiste à supprimer purement et simplement les observations manquantes. La méthode dite de « l'effacement de cas » (généralement nommée sous sa terminologie anglaise « *Listwise deletion* ») supprime ainsi l'ensemble des observations pour lesquelles au moins une valeur est non renseignée. Une variante, moins destructrice d'information, propose de supprimer uniquement les observations incomplètes lorsque celles-ci sont nécessaires à l'aboutissement des calculs (méthode des paires ou des « séries appariées », appelée « *Pairwise deletion* » en anglais). Ces premières méthodes ont le désavantage évident de réduire significativement la taille de l'échantillon d'observation. Par conséquent, elles ont très vite été supplantées par des techniques plus complexes

permettant de reconstruire les valeurs manquantes. En dépit du fait qu'une estimation inexacte de ces dernières peut induire un biais significatif sur les résultats des modèles (*Cf.* Harel, 2008), nous privilégions cette approche.

Dans le cas spécifique des bases de données financières, l'absence de certaines valeurs peut survenir pour de multiples raisons : erreurs du fournisseur, filtres préalables sur les valeurs aberrantes, *etc.* Plusieurs méthodes de simulations et de recouvrement de ces données sont proposées dans la littérature, mais celles-ci nécessitent généralement une connaissance *a priori* du type de données. En ce qui concerne l'estimation des données manquantes dans des séries temporelles de rendements d'actifs boursiers, l'algorithme de maximisation de la vraisemblance des valeurs manquantes (en anglais, « *Expectation-Maximization* » ; *Cf.* Dempster *et alii*, 1977) et ses nombreuses variantes (« *Expectation-Conditional-Maximization* » par exemple ; *Cf.* McLachlan et Krishnan, 1997) constitue un des modèles de référence. La simulation a de très nombreuses applications en finance (*Cf.* Jamshidian et Zhu, 1996). Boyle (1977), par exemple, propose de recourir aux simulations de Monte Carlo pour générer des trajectoires d'actifs et évaluer le prix d'options. Ce type de méthodes a largement gagné en popularité ces dernières années. Les générations de *scenarii* sont par exemple souvent réalisées *via* des méthodes de Monte Carlo (*Cf.* Guastaroba *et alii*, 2009). Celles-ci nécessitent toutefois d'effectuer des hypothèses sur la distribution des rendements, les lois normales ou de Student étant généralement retenues.

L'hypothèse de normalité des rentabilités des actifs est par ailleurs largement rejetée (*Cf.* Fama, 1965 ; Mandelbrot, 1997). De nombreuses études mettent en exergue les faits stylisés observés sur les rendements pour justifier le rejet du postulat Gaussien (*Cf.* Bollerslev *et alii*, 1992 ; Pagan, 1996 ; Harvey et Shephard, 1996 ; Campbell *et alii*, 1997 ; Cont, 2001 et Gouriéroux et Jasiak, 2001). Les rentabilités sont en effet asymétriques, leptokurtiques, hétéroscédastiques et caractérisées par une instabilité de leurs corrélations transversales (*Cf.* Gagliardini et Gouriéroux, 2006) vont également à l'encontre du postulat Gaussien). Toute modélisation Gaussienne tend donc à sous-estimer les risques extrêmes

(*Cf.* Alexiev, 2005). La dissymétrie qui peut être liée aux bulles spéculatives (*Cf.* Blanchard et Watson, 1982 ; Abreu et Brunnermeier, 2003), aux modes d'incitation optionnelle (*Cf.* Brennan, 1993 ; Bris *et alii*, 2007), à l'hétérogénéité des investisseurs (*Cf.* Hong et Stein, 1999 et 2003 ; Chen *et alii*, 2001) et aux risques de crédit et de liquidité (*Cf.* Kritzman, 1994) est par ailleurs renforcée pour les fonds alternatifs qui ont recours aux ventes à découvert et qui adoptent des stratégies dynamiques. Lorsque l'on considère les bases de données financières de fonds, *a fortiori* alternatifs, l'hypothèse de normalité ne peut en conséquence être retenue (*Cf.* Sfiridis, 2005). Les modèles de complétion de valeurs manquantes classiques doivent donc être adaptés aux spécificités des actifs financiers.

La nécessité de disposer d'un modèle d'estimation des valeurs manquantes est d'autant plus cruciale que les bases de données de fonds alternatifs sont biaisées compte tenu des pratiques en cours sur les marchés financiers. Par exemple, la possibilité de ne pas communiquer la performance des fonds est à l'origine de deux biais majeurs : celui d'auto-sélection et le biais dit « d'histoire instantanée ». Avant de présenter ces derniers, il nous faut rappeler que les fonds alternatifs sont soumis à des règles strictes concernant leurs moyens de communication. Le recours à la publicité leur est par exemple interdit et la communication de leurs performances passées pas toujours permise. Ainsi l'inclusion d'un fonds alternatif au sein d'une base de données offre un moyen légal de communication. Le choix d'intégrer leurs performances dans une base de données est généralement réalisé par des gestionnaires cherchant à attirer de nouveaux investisseurs. Le biais d'auto-sélection (en anglais, « *self-reporting* » ou « *superior funds* ») provient du fait que les fonds dont la taille est jugée optimale n'ont plus intérêt à communiquer leurs performances. Le biais « d'histoire instantanée » (en anglais, « *backfill* » ou « *instant history* » ; *Cf.* Fung et Hsieh, 2000 ; Lhabitant, 2004) résulte des pratiques des gestionnaires alternatifs qui attendent la fin de la période d'incubation (période consécutive à la création du fonds, qui correspond à la mise en place et la validation de la stratégie) pour inclure leur fonds au sein d'une base de données publique et reconnue. Lors de l'inclusion de leur fonds, ils ont la possibilité de reporter (une partie de) l'historique de performances passées correspon-

dant à la période d'incubation. Ainsi, seuls les gestionnaires qui réalisent des performances positives sur cette période vont choisir stratégiquement de renseigner celles-ci.

Plusieurs auteurs ont tentés d'estimer l'impact de cette pratique sur les bases de fonds et d'indices alternatifs (*Cf.* Brown *et alii*, 1999 ; Fung et Hsieh, 2000 ; Posthuma et van der Sluis, 2003). S'il n'est pas possible de connaître la part des fonds qui n'atteignent pas la phase de maturation, certains fournisseurs d'indices précisent néanmoins la date d'inclusion du fonds dans l'indice. Il est ainsi possible de dissocier les rendements ayant été rajoutés *ex post* des autres. L'impact du biais d'histoire instantanée est ainsi directement mesurable. Cet impact du biais d'histoire instantanée sur les rendements annualisés est estimé en moyenne entre 2 et 6 points suivant les études. En dépit de certaines divergences concernant l'estimation du biais, toutes les études s'accordent néanmoins sur le fait que les performances passées sont surestimées et parfois lissées. Ainsi, Fung et Hsieh (2000), Posthuma et van der Sluis (2003) et Malkiel et Saha (2005) recommandent de supprimer systématiquement les rendements correspondants à la période d'incubation (comprise en général entre un et deux ans d'après ces mêmes auteurs). La suppression de ces périodes d'incubation ainsi que la nécessité de disposer de base de données cylindrée rend ainsi nécessaire de disposer de modèles de complétion de valeurs manquantes. En outre, la méthode de reconstruction retenue doit être adaptée aux spécificités des rendements empiriques, ceux des fonds alternatifs en particulier. Le premier chapitre tente de répondre à cette problématique. Dans une première partie nous présentons notre modèle de traitement des valeurs manquantes. Celui-ci résulte de la combinaison de trois méthodes : les cartes de Kohonen (*Cf.* Kohonen, 1995), les fonctions orthogonales empiriques (*Cf.* Hanson et Lawson, 1974) et l'algorithme des « *Generalized Surrogate Data* » (*Cf.* Schreiber, 1998). La problématique de génération de *scenarii* est ensuite abordée. Des *scenarii* réalistes sont obtenus en tirant partie des propriétés d'interpolation non-linéaire des cartes de Kohonen. Cette approche permet la génération de *scenarii* sans avoir à spécifier d'hypothèse forte sur les distributions des rendements des actifs à simuler.

Les cartes de Kohonen (ou cartes auto-organisées, notées *SOM* pour « *Self-Organizing Maps* » en anglais) constituent un puissant outil de classifications issues des techniques neuronales. Les *SOM* correspondent à une version stochastique de l'algorithme des Centres Mobiles (également appelé « algorithme de Forgy » ; *Cf.* Forgy, 1965). Au moyen d'une projection non-linéaire sur un espace de faible dimension (ficelle ou grille), les *SOM* permettent simultanément d'explorer, de regrouper et de visualiser des données de grandes dimensions. Elles sont utilisées dans de nombreux domaines, et en finance en particulier (*Cf.* Deboeck et Kohonen, 1998 ; Maillet et Rousset, 2003 ; Ben Omrane et de Bodt, 2007). Au moyen d'un apprentissage itératif, les cartes de Kohonen permettent de générer des groupes homogènes d'individus, ainsi que des individus représentatifs (également appelés « vecteurs codes ») de chacun de ces groupes. Ce type de réseau a de bonnes propriétés de convergence (*Cf.* Cottrell et Fort, 1987), et ce malgré l'absence de certaines valeurs dans la base d'apprentissage (*Cf.* Samad et Harp, 1992). Une simple adaptation de l'algorithme classique rend possible l'estimation des valeurs manquantes (*Cf.* Fessant et Midenet, 2002 ; Wang, 2003). Ainsi, lorsqu'une base de séries temporelles n'est pas complète, les *SOM* permettent certes de classifier les séries, mais aussi de générer les individus représentatifs de chacun des groupes : les vecteurs codes. Ces vecteurs ne contiennent, par définition, aucune valeur manquante. Il est ainsi assez naturel d'estimer les valeurs manquantes des séries à l'aide de celles correspondantes à l'individu représentatif du groupe auquel appartient la série. Les *SOM* permettent ainsi très simplement de reconstruire et d'estimer, de façon non-linéaire, les valeurs non renseignées. Néanmoins, de par leur construction, les individus représentatifs s'apparentent en quelque sorte à des « moyennes pondérées » des séries de la base d'apprentissage. Le remplacement des valeurs manquantes par celles des vecteurs codes correspondants a pour conséquence d'engendrer des séries dont la volatilité est sous-estimée. De plus, cette estimation est discrète dans la mesure où lorsque deux séries appartiennent à un même groupe et ont des valeurs non renseignées pour une même observation (date), elles se verront attribuer un même candidat pour le remplacement de ces valeurs manquantes. Ainsi, cette reconstruction ne permet pas de restituer fidèlement les moments d'ordre supérieur ou égal à deux. En outre, les structures de dépendance

des séries appartenant à un même groupe seront également modifiées. Un autre algorithme permettant la reconstruction des valeurs manquantes est obtenu en considérant les Fonctions Orthogonales Empiriques (en anglais *Empirical Orthogonal Functions*, notées *EOF*). Les *EOF* constituent une méthode de décomposition factorielle initialement utilisée en traitement du signal. Elles permettent la décomposition d'un ensemble de données en fonction de bases orthogonales. Cet algorithme s'apparente à celui, plus connu, de l'Analyse en Composante Principale (noté *PCA* pour *Principal Component Analysis* en anglais). La méthode des *EOF* se différencie de celle de la *PCA* dans la mesure où que les composantes sont déterminées non seulement à partir des individus (séries temporelles), mais (à la différence de la *PCA*) également en fonction des observations (dates). De façon analogue à la *PCA*, les *EOF* permettent d'appliquer un processus de débruitage des données en supprimant la partie résiduelle des séries correspondant aux composantes négligeables obtenues. Les *EOF* permettent de reconstruire des données de façon continue et linéaire. Après une première initialisation des valeurs manquantes (moyenne des séries par exemple), l'application de la technique de « débruitage » par les *EOF* permet d'obtenir un nouvel ensemble de candidats, plus vraisemblables, pour le remplacement des valeurs manquantes. L'application successive de cette technique permet de converger vers des estimations plus précises de ces valeurs. Cette méthode, initialement proposée pour la reconstruction des images satellitaires (*Cf.* Boyd *et alii*, 1994), a ensuite été appliquée à de nombreux types de données (*Cf.* van den Branden et Verboven, 2009). La qualité des estimations obtenues *via* les *EOF* est largement conditionnée par le choix des premières initialisations. La qualit des reconstructions issues des *EOF* est de plus limitée par le fait que celles-ci sont linéaires.

Les *SOM* et les *EOF* constituent donc deux techniques de reconstruction des valeurs manquantes qui sont complémentaires. La première méthode est non-linéaire mais discrète, alors que la seconde est linéaire mais continue. Nous proposons ainsi une combinaison de ces deux méthodes permettant de tirer partie des avantages de chacune. Les premières estimations issues des cartes de Kohonen servent à initialiser la seconde méthode. Enfin, nous proposons de compléter l'algorithme de complétion combiné, par un algorithme issu de la

physique appliquée appelé *Generalized Surrogate Data Method* (*Cf.* Schreiber, 1998). Cet algorithme itératif, inspiré du recuit simulé (également appelé algorithme de Metropolis-Hastings ; *Cf.* Chib et Greenberg, 1995) permet la génération de nombres aléatoires respectant des contraintes prédéfinies. Ainsi, la méthode des *Generalized Surrogate Data* appliquée à nos estimations de valeurs manquantes nous permet de nous assurer que certaines particularités statistiques (tels les niveaux de corrélation ou les moments d'ordre supérieur par exemple ; *Cf.* Chen *et alii*, 2008) des séries temporelles reconstruites sont bien respectées.

Les problèmes rencontrés lors de la génération de valeurs réalistes nous conduisent naturellement à nous intéresser aux modèles de génération de *scenarii* (*Cf.* Jamshidian et Zhu, 1996). Un *scenario* correspond à la réalisation d'un ensemble de variables multivariées (des rendements d'actifs dans notre cas). Plusieurs solutions d'implémentation regroupent ce type de simulations : de la simple approche historique, à des méthodes plus complexes utilisant des techniques de ré-échantillonnage des observations passées (les méthodes de *Bootstrap* ; *Cf.* Efron et Tibshirani, 1993), des méthodes de simulations factorielles (*Cf.* Loretan, 1997) ou encore des méthodes de simulations paramétriques de données multivariées (simulations de Monte Carlo ; *Cf.* Glasserman, 2003).

L'approche historique est une méthode particulièrement répandue, probablement en raison de sa simplicité. Elle ne nécessite que de faibles hypothèses concernant les distributions des rendements ou les structures de dépendance entre les actifs considérés. Ces derniers sont implicitement intégrés à partir des observations passées. Cette approche nécessite de disposer de séries historiques de prix suffisamment longues. En effet, une simulation historique qui serait fondée sur une période d'analyse trop courte (ou non représentative) génèrerait des données biaisées. De plus, le nombre de *scenarii* réalisables est limité par le nombre d'observations disponibles. Les techniques de *scenarii* ré-échantillonnés combinent les méthodes de simulations historiques et celles, usuelles, de ré-échantillonnages. A partir des données historiques, chaque réalisation correspond à un *scenario*. Les *scenarii* sont

ensuite ré-échantillonnés. Le *bootstrap* rend possible la simulation de *scenarii* lorsque le nombre de réalisations est faible. Cette technique conserve les avantages de l'approche historique (*Cf.* Kouwenberg, 2001). Elle est toutefois sujette à des bais (*Cf.* Kohavi, 1995) et nécessite que l'horizon des *scenarii* soit supérieur à la périodicité de la base de données d'entrée. Puisque les séries de fonds alternatifs ont généralement une périodicité mensuelle, il sera impossible d'obtenir des *scenarii* d'horizon inférieur ou égal à un mois.

Enfin, les méthodes de simulations reposent sur diverses hypothèses qui caractérisent les processus régissant les rendements des actifs. Les simulations de Monte Carlo consistent à générer des réalisations de rendements après en avoir spécifié leurs distributions. Le choix des distributions adéquates est effectué au regard de tests statistiques réalisés sur des données empiriques (des séries de rendements). Ainsi, lorsque l'on considère des *scenarii* définis sur plusieurs actifs, l'estimation de la distribution jointe peut s'avérer complexe. En outre, la puissance des tests d'adéquation des modèles est restreinte dès que l'on simule plusieurs séries. L'hypothèse Gaussienne permet de contourner ces difficultés. En effet, les distributions marginales (correspondant à un actif) sont obtenues par l'estimation des deux premiers moments empiriques et la structure de dépendance est complètement définie à partir de la matrice de corrélation. Lorsque l'hypothèse de normalité est écartée, la fonction de copule permet de définir explicitement la dépendance entre les variables (*Cf.* Nelson, 1998). Les copules permettent ainsi de construire une variable multivariée à partir des différentes lois marginales. Leur considération permet ainsi de répondre à de nombreuses problématiques financières (modèles de gestion des risques et de simulation ; *Cf.* Jouanin *et alii*, 2004 ; Kaut et Wallace, 2006). L'approche par les copules est néanmoins difficilement envisageable dès lors qu'un nombre important d'actifs est considéré (*Cf.* Mikosch, 2006). D'autres techniques de simulations s'attachent à modéliser le plus fidèlement possible certains faits stylisés concernant les rendements. Par exemple, les processus de type autorégressif (du type « *Multivariate Generalized AutoRegressive Conditional Heteroskedastic* » en anglais, noté *M-GARCH* ; *Cf.* Bollerslev *et alii*, 1988) modélisant la variance conditionnelle, permettent de rendre compte des effets de la corrélation sérielle des volatilités. La calibration de ce type de modèle se fonde sur les réalisations passées conjointes

des rendements et de la volatilité. Ce type d'approche n'est ainsi pas approprié pour des fonds alternatifs car les performances peuvent être manipulées (*Cf.* Goetzmann *et alii*, 2007) et la volatilité ainsi sous-évaluée et/ou lissée.
Nous écartons donc les approches paramétriques et nous retenons le cadre du ré-échantillonnage historique des rendements pour développer une méthode de génération de *scenarii*. Nous exploitons à nouveau les caractéristiques des cartes de Kohonen afin d'établir une classification transversale des séries temporelles. Cette classification permet d'obtenir des vecteurs codes caractéristiques d'une réalisation multivariée, en d'autres termes un *scenario*. A chaque neurone de la carte va correspondre un *scenario*. En exploitant les propriétés d'interpolation des *SOM* (*Cf.* Sarzeaud et Stéphan, 2000), la création de cartes de très grandes tailles autorise la génération d'un nombre important de *scenarii*. Notre modèle conserve les bonnes propriétés des modèles de simulations historiques (aucune hypothèse forte sur la distribution des rendements n'est nécessaire). En outre, il autorise une modélisation plus fine des *scenarii*. Nous l'illustrons en comparant nos résultats à ceux obtenus *via* un ré-échantillonnage historique et les simulations de Monte Carlo retenant des hypothèses de distributions normale et de Student. Nos résultats suggèrent que notre modèle semble bien adapté à la génération de *scenarii*. En supposant que les réalisations simulées sont équiprobables, il nous est alors possible d'en déduire un ensemble de mesures de risque.

Après une présentation détaillée des méthodes proposées, nous les appliquons en considérant un ensemble de séries de rendements de fonds alternatifs issu de la base de données HFR^{TM}. L'ensemble choisi est constitué des rendements mensuels en dollars américains de 50 fonds alternatifs sur une période allant de décembre 1995 à décembre 2005. La base des fonds retenus est initialement sans aucune valeur manquante. Enfin, les fonds considérés suivent des stratégies diverses. Cette diversité induit notamment l'existence de dépendances non-linéaires. Elle rend ainsi l'exercice de simulation particulièrement intéressant.

Afin d'évaluer la qualité de notre méthode de reconstruction des valeurs manquantes, nous détruisons artificiellement et aléatoirement une partie des observations de la base de données. Différentes proportions de valeurs manquantes (de 5% à 25% des observations) sont considérées. Nous proposons d'évaluer la qualité des reconstructions fournies par notre modèle en les comparant avec celles issues d'algorithmes standard (*ECM* dans notre cas). De même, les *scenarii* générés à partir de cette même base par notre modèle de classification transversale sont comparés à ceux obtenus par simulation de Monte Carlo (avec des hypothèses distributionnelles normale et de Student).

Nous montrons enfin l'impact de ces reconstructions et simulations sur les résultats de modèles financiers répandus. Plus précisément, nous évaluons les conséquences de l'application de méthodes de reconstruction inadaptées sur les compositions des portefeuilles obtenus par des modèles d'allocation d'actifs classiques, ainsi que dans le cadre du calcul de mesures de risque traditionnelles (Valeur-en-Risque). Ces premières expérimentations semblent valider notre modèle d'estimation de valeurs manquantes. Sur cette base de fonds alternatifs, les reconstructions et simulations sont en effet en moyenne meilleures que celles obtenues *via* les méthodes *ECM* ou Monte Carlo.

Les premiers résultats obtenus, tout à fait encourageants, nous incitent à prolonger cet axe de recherche. Il apparaît ainsi nécessaire d'appliquer ces méthodes à différentes bases de données (notamment sur des bases constituées d'actifs dont les rentabilités sont clairement non-normales) ainsi qu'à des données simulées. L'ajout de contraintes sur les caractéristiques des séries reconstruites (l'autocorrélation notamment) pourra améliorer sensiblement notre méthode.

Le modèle de simulation proposé ouvre également de nombreuses perspectives. Il sera tout d'abord nécessaire de présenter un ensemble de tests sur les structures des dépendances des séries reconstruites ou simulées ainsi que sur la fiabilité des mesures de risques obtenues, et notamment des Valeur-en-Risques, lorsque les données sont complétées ou simulées (*Cf.* Kupiec, 1995 ; Christoffersen, 1998 ; Engle et Manganelli, 2004). Si les dépendances des séries simulées reflètent bien celles des séries initiales, nous pourrons

alors proposer des améliorations pour l'ensemble des tests de détermination de copule.
Il sera également intéressant de développer ce dernier afin de prendre en compte les probabilités d'occurrence des *scenarii*. La considération de mesure d'intensité de crise (par exemple, « *Index of Market Shocks* », *Cf.* Maillet et Michel, 2003) permettra en outre d'adopter une approche conditionnelle. Il sera possible de déterminer les états des marchés auxquels correspondent les *scenarii* générés. Nous pourrons ainsi considérer uniquement les *scenarii* correspondant à un état de crise (*Cf.* Annexe 5), et en déduire des mesures de risques extrêmes (de type *stress test*).

Résumé du chapitre sur "*Efficient Frontier for Robust Higher-moment Portfolio Selection*"

L'hypothèse d'investisseurs averses au risque maximisant l'espérance d'utilité de leur richesse future, proposée initialement par Bernoulli (1738) et développée par von Neumann et Morgenstern (1944), est au cœur de la théorie financière moderne. Associée à l'hypothèse de non-satiété, l'aversion au risque des investisseurs implique des fonctions d'utilité strictement croissantes et strictement concaves. Sous certaines conditions, ce cadre « standard » d'analyse d'agents averses au risque maximisant leur espérance d'utilité peut se ramener au cadre « espérance-variance » proposé par Markowitz (1952). En considérant des fonctions d'utilité quadratiques (*Cf.* Fishburn, 1979) ou des distributions elliptiques symétriques – à laquelle appartient la loi normale – (*Cf.* Chamberlain, 1983 ; Owen et Rabinovitch, 1983), la préférence d'un agent pour un investissement est entièrement déterminée par les deux premiers moments de la distribution des rentabilités.

Il est toutefois aujourd'hui bien connu que ces deux types d'hypothèses apparaissent fragiles. D'une part, une fonction de préférence quadratique implique la propriété indésirable de satiété (l'augmentation de la richesse réduit l'utilité) pour un certain niveau de richesse, ainsi qu'une aversion absolue pour le risque croissante (la demande d'actifs risqués diminue lorsque la richesse s'accroît), ce qui revient à admettre que les actifs financiers sont des biens « inférieurs » (*Cf.* Arrow, 1964 ; Pratt, 1964 ; Huang et Litzenberger, 1988). Par ailleurs, l'hypothèse d'une fonction de préférence quadratique ne permet pas de rendre compte de certains paradoxes. Par exemple, dans un tel cadre, la participation d'agents à des loteries risquées ne peut être justifiée (*Cf.* Friedman et Savage, 1948 ; Kahneman et Tversky, 1979 ; Golec et Tamarkin, 1998 ; Barberis et Huang, 2007) et les comportements des agents tels que la prudence et la tempérance (*Cf.* Kimball, 1990 et 1993) ne peuvent pas être pris en compte. D'autre part, l'hypothèse de distributions des rentabilités elliptiques symétriques pose elle aussi problème car elle ne permet pas de modéliser le caractère asymétrique et leptokurtique des rentabilités des actifs financiers (*Cf.* Fama, 1965 ; Bollerslev, 1987 ; Jondeau et Rockinger, 2003-a). Comme rappelé lors

du chapitre précédent, l'asymétrie de la distribution des rentabilités des actifs peut s'expliquer par l'existence de phénomènes de bulles spéculatives (*Cf.* Blanchard et Watson, 1982 ; Abreu et Brunnermeier, 2003) ou de modes d'incitation optionnels (*Cf.* Black et Scholes, 1973 ; Brennan, 1993 ; Bris *et alii*, 2007). Plus généralement, les phénomènes de contagion et d'hétéroscédasticité des rentabilités, ainsi que l'hétérogénéité des anticipations des agents (*Cf.* King et Wadhwani, 1990 ; Hamao *et alii*, 1990 ; Embrechts *et alii*, 1999 ; Lillo *et alii*, 2009), sont à l'origine du caractère asymétrique et leptokurtique des fonctions de densité des rentabilités des actifs. Les moments d'ordre supérieur à celui de la variance ont ainsi été introduits dans l'analyse afin de rendre compte plus finement des caractéristiques des préférences des agents (*Cf.* Marschak, 1938 ; Hicks, 1939). Leur importance a longtemps été justifiée par le fait que ces moments apparaissent explicitement dans les développements limités des fonctions d'utilité. Ainsi, la dérivée d'un développement de Taylor d'une fonction d'utilité d'un agent globalement averse au risque par rapport à la *skewness* est généralement positive (*Cf.* Arditti, 1967 ; Kraus et Litzenberger, 1976). Pour autant, Brockett et Kahane (1992) et Brockett et Garven (1998) montrent qu'il n'est pas possible d'établir une relation systématique entre le signe des dérivés d'une fonction d'utilité et les préférences pour les moments.
Initialement proposée par Arditti (1967), Levy (1969), Jean (1971), Rubinstein (1973), Ingersoll (1975), et Kraus et Litzenberger (1976), pour rendre compte de l'influence du coefficient d'asymétrie de la distribution inconditionnelle des rentabilités boursières sur les préférences des agents et l'évaluation des actifs, l'analyse des moments d'ordre supérieur a ensuite été étendue aux quatre premiers moments inconditionnels par Homaifar et Graddy (1988), Athayde et Flôres (1999) et Hwang et Satchell (1999).

La non-normalité des rendements des actifs d'une part et les préférences des agents d'autre part (*Cf.* Scott et Horwath, 1980) imposent de prendre en considération les caractéristiques de dissymétrie et de queues de distributions épaisses dans les modèles de choix de portefeuille. Nous proposons, dans ce chapitre, une approche non-paramétrique de la frontière efficiente traduisant complètement, en termes de composition de porte-

feuille, tous les arbitrages possibles entre les différents moments de la distribution qui déterminent la fonction d'utilité. Nous supposons que le comportement des investisseurs peut être décrit par des fonctions d'utilité compatibles avec le critère de dominance stochastique d'ordre quatre (*Cf.* Levy, 1992 ; Vinod, 2004), ce qui nous permet de lier le critère de l'espérance d'utilité à l'ensemble des moments de la distribution des rentabilités par l'utilisation d'un développement en série de Taylor. L'introduction d'une fonction d'utilité quartique nous permet ensuite d'obtenir un critère de décision en incertitude dans un cadre moyenne-variance-asymétrie-*kurtosis* exact (*Cf.* Benishay, 1987, 1989 et 1992). Ce type de fonction d'utilité peut en effet satisfaire – pour un choix judicieux de paramètres – les cinq propriétés désirables de tout investisseur rationnel (*Cf.* Pratt, 1964 ; Arrow, 1970 ; Kimball, 1990) : à savoir la "gourmandise", la stricte aversion pour le risque, un coefficient d'aversion absolue pour le risque décroissant en la rentabilité, un coefficient de prudence absolue strictement décroissant avec la rentabilité, et enfin, une aversion relative pour le risque constante ou croissante en la rentabilité (*Cf.* Crainich et Eeckhoudt, 2008).

Deux approches sont généralement distinguées dans la littérature pour construire l'ensemble des portefeuilles optimaux au sens du critère de décision moyenne-variance-asymétrie-*kurtosis*. La première, dite primale, s'attache à déterminer l'ensemble de portefeuilles telles que leurs caractéristiques en termes de moments soient optimales. La seconde, duale, s'attache à maximiser une fonction indirecte d'utilité qui dépend des moments de la distribution des rentabilités associées à l'investissement.

La première approche est mobilisée, par exemple, par Berényi (2001 et 2002) et Davies *et alii* (2006). Ils définissent un programme de maximisation sous contraintes pour déterminer l'ensemble des fonds de fonds alternatifs efficients au sens du critère moyenne-variance-asymétrie-*kurtosis*. Leur modèle ne permet cependant pas d'établir de relation explicite avec la satisfaction de l'agent car les paramètres pondérant chacun des moments optimisés ne sont pas directement liés à l'espérance d'utilité de l'investisseur. Par ailleurs, l'allocation optimale proposée, issue du programme de maximisation, n'est au final qu'une approximation, c'est-à-dire un compromis, puisque ce programme consiste à minimiser les

déviations par rapport aux autres moments. Cette démarche ne permet d'obtenir qu'une solution « proche » de la frontière efficiente élargie, mais la solution proposée n'est pas nécessairement optimale au sens de Pareto. Dans la même catégorie d'approches, certains auteurs s'attachent à résoudre analytiquement le programme d'optimisation. Par exemple, Athayde et Flôres (1999), Adcock (2004 et 2008), Jurczenko et Maillet (2006-b) ainsi que Mencia et Sentana (2008), proposent des solutions analytiques caractérisant la frontière efficiente étendue, avec l'objectif de minimiser la variance pour une espérance, une asymétrie et un indice de *kurtosis* donnés. Ces approches analytiques souffrent également du caractère restrictif des liens supposés entre les moments (*Cf.* Athayde et Flôres, 2003) et de l'inadéquation avec les données réelles de l'hypothèse distributionnelle (*Cf.* Adcock, 2008). Elles restent également partielles dans leurs objectifs puisqu'elles s'intéressent principalement au moment d'ordre deux au détriment des autres. L'approche primale du choix de portefeuille est ainsi viable dans le plan restreint de l'espérance-variance, dans la mesure où l'univers des portefeuilles (et leurs combinaisons) constitue un ensemble convexe. Une relation d'ordre (de préférence) peut ainsi être définie. Il y a, en outre, un arbitrage entre le rendement attendu d'un portefeuille et le risque associé. L'élargissement de l'approche primale aux moments d'ordre supérieur apparaît cependant particulièrement délicate car une relation d'ordre ne peut être définie (*Cf.* Zhang, 2008). Dans le cadre étendu aux quatre premiers moments, les combinaisons des portefeuilles décrivent un ensemble non nécessairement convexe. La considération du problème dual devient ainsi nécessaire pour déterminer les préférences des investisseurs puisqu'il est impossible de définir une relation d'ordre dans le cadre primal.

L'approche duale a pour origine la spécification d'une fonction d'utilité indirecte et elle consiste en une expansion de Taylor (*Cf.* Loistl, 1976) de la fonction objective du programme, pour déterminer les portefeuilles optimaux (*Cf.* par exemple Jondeau et Rockinger, 2003-a et 2006 ; Jurczenko et Maillet, 2006-b). Cette approche est courante dans le cadre des études empiriques des modèles multi-moments. En effet, l'introduction de fonctions d'utilité spécifiques induit une agrégation des préférences des investisseurs pour les moments et elle permet de réduire largement la complexité du programme de

maximisation de l'utilité du portefeuille. Outre la nécessité de choisir *a priori* le type de la fonction d'utilité, le principal inconvénient de cette approche concerne l'application d'une expansion de Taylor qui ne converge vers l'utilité espérée que sous certaines conditions sur les rendements (*Cf.* Loistl, 1976 ; Jondeau et Rockinger, 2003 ; Jurczenko et Maillet, 2006-a ; Garlappi et Skoulakis, 2008). Pour certaines formes de fonction d'utilité (exponentielle par exemple), cette convergence n'est de plus assurée que pour des plages restreintes de réalisation des rendements, ce qui reste problématique en finance (spécifiquement pour quelques variétés d'investissements) en raison de la fréquence importante des larges déviations. Par ailleurs, la prise en compte de moments supérieurs, lorsqu'ils existent, ne conduit pas nécessairement à une amélioration de l'approximation (*Cf.* Jurczenko et Maillet, 2006-b).

Toutefois, puisque la frontière efficiente élargie est une surface non convexe, les approches duales – tout autant que les primales – permettent d'obtenir uniquement des solutions optimales locales, mais pas nécessairement globales. Il est en effet nécessaire de rendre convexe le problème en ajoutant des restrictions sur les moments, sur les fonctions de densité (séparantes) ou encore sur les fonctions d'utilité – voir Rubinstein (1973), Ingersoll (1975) et Athayde et Flôres (2004). En particulier, l'approche duale spécifique implique le risque de ne pas obtenir de solution atteignable compatible avec la technologie du marché. De plus, lorsque la dimension du problème (le nombre d'actifs) croît, il devient particulièrement complexe d'interpréter géométriquement la frontière efficiente. La détermination des portefeuilles optimaux n'est ainsi pas assurée.
Pour contourner en partie ces différentes limites, nous retenons une méthode de résolution de problème à objectifs multiples : les modèles de *Polynomial Goal Programming* (*PGP*). Cette approche rend ainsi possible l'optimisation simultanée des quatre premiers moments. La version standard du *PGP*, développée par Charnes et Cooper (1961), est un outil d'aide à la décision fondé sur une philosophie de satisfaction globale qui a pour rôle d'assister le décideur cherchant à satisfaire simultanément différents objectifs et d'évoluer vers une solution de compromis satisfaisante. Dans sa version initiale, le modèle du *PGP* ne permet

pas de prendre en compte les interactions non-linéaires entre les contraintes. A l'instar de Briec *et alii* (2004 et 2007), nous choisissons alors d'utiliser la fonction de distance, également appelée fonction « de pénurie » (*shortage function* en anglais). Cette fonction de pénurie, introduite à l'origine en théorie de la production par Luenberger (1995), est une fonction de distance qui combine simultanément réduction d'*inputs* et augmentation des *outputs*, dont la fonction duale est celle du profit de l'entreprise. Elle offre une parfaite représentation d'un ensemble de choix multidimensionnels et elle permet d'évaluer la position de toute combinaison d'*intputs-outputs* par rapport à la frontière de l'ensemble des portefeuilles possibles. Elle a de plus l'avantage d'intégrer les deux approches primale et duale. Cette méthodologie a déjà été appliquée en finance par Morey et Morey (1999) et Briec *et alii* (2004) pour évaluer la performance des fonds et par Briec *et alii* (2007) pour obtenir des portefeuilles optimaux avec des asymétries sous-jacentes significatives. La fonction de pénurie permet ainsi de projeter n'importe quel portefeuille (*a priori*) inefficient, sur la frontière efficiente élargie. Elle permet de juger de la performance d'un portefeuille par rapport à sa projection sur la frontière efficiente primale.

Suivant l'approche de Briec *et alii* (2004 et 2007), nous montrons que la fonction de pénurie permet d'obtenir un *optimum* global. Nous établissons l'équivalence duale entre notre approche et la fonction d'utilité indirecte représentée par les moments des distributions des rentabilités possibles. Ces deux résultats nous permettent de conclure que notre approche domine les précédentes approches duale et primale proposées dans la littérature. Elle ne suppose pas en outre l'existence d'un actif sans risque et elle permet aisément de restreindre strictement les ventes à découvert.

Nous généralisons les résultats de Briec *et alii* (2004 et 2007) en considérant le cadre d'analyse espérance-variance-asymétrie-*kurtosis*. Puisque la fonction de pénurie dépend des quatre premiers moments d'un portefeuille, il est nécessaire d'expliciter les moments des rendements du portefeuille par les co-moments associés des rentabilités des actifs considérés. Athayde et Flôres (2002) ont proposé une notation compacte, utilisant les produits tensoriels (des produits de Kronecker) et des matrices de co-moments, pour ca-

ractériser les moments d'ordre trois et quatre. Nous proposons une nouvelle formulation des moments d'ordre supérieur, parfaitement équivalente, mais permettant de réduire sensiblement le temps de calcul des différents moments. Des vecteurs de co-moments sont introduits. Quel que soit l'ordre considéré, une formulation récursive nous permet alors de définir ces vecteurs.

Nous utilisons également des statistiques robustes pour l'estimation des moments. En effet, une des principales limites des applications du cadre moyenne-variance-asymétrie-*kurtosis* réside dans le problème d'estimation des moments. Les biais survenant lors de l'estimation de la moyenne et la variance impactent en effet significativement les portefeuilles optimaux (*Cf.* Michaud, 1989 ; Best et Grauer, 1991). Les coefficients d'asymétrie et d'aplatissement sont plus sensibles encore aux données aberrantes (*Cf.* Kim et White, 2004). Ainsi la puissance explicative ou descriptive des modèles d'allocation d'actifs incorporant des moments d'ordre supérieur est limitée en raison de ces difficultés à estimer fidèlement les moments. Nous proposons ici de développer la technique de sélection de portefeuilles proposée dans Jurczenko *et alii* (2006), en substituant aux moments traditionnels des statistiques robustes : les *L-moments*.

Introduits par Sillitto (1951) et popularisés par Hosking (1989), les *L-moments* permettent de caractériser, comme les moments classiques, la forme des distributions statistiques. Ils correspondent à des combinaisons linéaires de statistiques d'ordre. Outre leur robustesse, ils présentent de nombreux avantages par rapport aux moments traditionnels, spécifiquement dans les champs de la finance empirique. D'abord, dès lors que la moyenne existe, tous les *L-moments* existent (*Cf.* Chan, 1967 ; Arnold et Meelden, 1975). Contrairement aux moments traditionnels, une distribution peut toujours être entièrement caractérisée par ses *L-moments*, même si certains de ses moments classiques ne sont pas définis (*Cf.* Jondeau et Rockinger, 2003-a ; Jurczenko et Maillet 2006-a). Hosking et Wallis (1987) privilégient ainsi l'utilisation des *L-moments* pour leur robustesse et parce qu'ils offrent, en comparaison avec les moments traditionnels, une meilleure approximation des lois de distribution en général. Les *L-moments* sont en particulier plus adaptés à l'estima-

tion des queues de distribution (valeurs extrêmes). En outre, les *L-moments* apparaissent particulièrement intéressants pour les applications financières, dans la mesure où ce sont des mesures de risque cohérentes (*Cf.* Artzner *et alii*, 1999). Enfin, les *L-moments* empiriques sont (plus) robustes aux données aberrantes (Vogel et Fennessey, 1993), puisqu'ils ne sont influencés que linéairement par les déviations extrêmes (*Cf.* Hosking, 1990) ; ils sont ainsi plus adéquats que les moments traditionnels dans le cadre de notre problématique (*Cf.* Sankarasubramanian et Srinivasan, 1999 ; Carrillo *et alii*, 2006-b).

Une illustration de notre modèle de sélection de portefeuille est ensuite proposée. La base retenue est constituée des cours d'actions européennes en euros fournis par *Bloomberg*. Nous choisissons les actions les plus liquides intégrées à l'indice *DJ European Stoxx* et nous retenons uniquement les titres des sociétés intégrés à l'indice sur l'ensemble de la période de juin 2001 à juin 2006. Enfin, les actions pour lesquelles des opérations sur titre (telles que des divisions d'action par exemple) induisent des variations de prix ne reflétant pas la variation de la valeur des sociétés sont écartées. Nous retenons finalement 162 titres des 600 initialement présents dans l'indice. Ces 162 titres représentent une capitalisation boursière flottante de plus de 1 300 milliards d'euros (soit près d'un quart de la capitalisation flottante de l'indice). La non-normalité des rendements des actifs présents dans cette base rend *a priori* l'application de notre modèle de sélection de portefeuille intéressante au regard des préférences des agents pour les moments d'ordre supérieur. Nous déterminons, au moyen de la fonction de pénurie, l'ensemble des portefeuilles optimaux dans l'espace moyenne-variance-asymétrie-*kurtosis*.

A partir de l'ensemble de ces portefeuilles optimaux, nous projetons les portefeuilles dans les plans correspondants à chacune des couples de moments. Nous évaluons ainsi les arbitrages entre les moments. Dans sa version robuste, la frontière efficiente, dans le plan des deux premiers moments, décrit une forme comparable à celle issue de l'analyse classique. Les investisseurs cherchant à accroître leur espérance de gains doivent accepter de supporter une plus grande dispersion des rendements. Nous observons également un lien fort entre les deuxième et quatrième moments. Les investisseurs cherchant à accroître leur

espérance de gains doivent également accepter un risque de perte extrême plus important. Ces observations, assez naturelles et intuitives, sont cohérentes avec celles de Maringer et Parpas (2009).

Les conclusions concernant le troisième moment sont moins évidentes. Nous notons, tout de même, que les portefeuilles qui ont les plus grandes espérances de rendement pour une volatilité minimum ont une distribution asymétrique à gauche (*skewness* négative). Nous observons ainsi le phénomène de prime pour l'asymétrie (*Cf.* Post *et alii*, 2008). Globalement, l'utilisation des *L-moments* conduit aux mêmes conclusions que celles obtenues à partir des moments traditionnels. La tempérance des investisseurs (mesurée par le *L-moment* d'ordre quatre des rendements des portefeuilles choisis) est étroitement reliée à leur aversion au risque (mesurée par le *L-moment* d'ordre deux). La relation entre les moments d'ordre pair et impair n'est toutefois pas linéaire. Ces premières conclusions suggèrent que l'asymétrie d'un portefeuille est un élément à prendre en considération lors de la construction d'un portefeuille efficient, alors que la *kurtosis*, très liée à la variance, peut être négligée.

Après avoir rappelé la formulation du problème de choix de portefeuille d'un agent rationnel cherchant à maximiser son utilité dans le cadre moyenne-variance-asymétrie-*kurtosis*, nous choisissons une classe de fonctions d'utilité très générale, reflétant une aversion au risque « mixte » et respectant les critères de dominance stochastique d'ordre quatre (*Cf.* Caballé et Pomansky, 1996). Cette classe de fonction d'utilité nous permet alors d'estimer le niveau d'utilité des agents au moyen d'une approximation de Taylor à l'ordre quatre. Les préférences des agents pour les moments d'ordre supérieur sont ainsi directement prises en compte. Nous considérons des fonctions d'utilité quartiques, *CRRA* (pour *Constant Relative Risk Aversion*) et des fonctions plus générales dites puissance-exponentielles. Considérant successivement les moments classiques et les *L-moments*, nous déterminons parmi les portefeuilles primaux optimaux, ceux qui le sont également dans le problème dual (pour différentes valeurs de paramètres). Quelle que soit la méthode d'estimation des moments (robuste ou traditionnelle), nous obtenons des portefeuilles très

proches de ceux de la frontière efficiente classique. Nous concluons ainsi comme Jondeau et Rockinger (2003-b) que les moments d'ordre supérieur sont à considérer dans le cas où les agents seraient particulièrement sensibles à ceux-ci, ou lorsque les données sont fortement non Gaussiennes. De plus, les *L-moments* permettent une réduction des problèmes d'estimation des moments traditionnels et ils sont ainsi à privilégier lors de la recherche de portefeuilles optimaux (*Cf.* Darolles *et alii*, 2008).

Ces derniers résultats nous encouragent à poursuivre dans le futur cette recherche en considérant des actifs dont les rentabilités s'écartent significativement de la rentabilité normale hypothétique. Il serait également intéressant de poursuivre l'étude de Davies *et alii* (2006) concernant la prise en compte de l'asymétrie et des risques extrêmes pour des modèles d'allocation de fonds de fonds alternatifs, caractérisés par des distributions s'éloignant très significativement de la distribution normale.

Un développement mathématique de la formulation du problème d'optimisation de portefeuille serait en outre intéressant. En effet, la nouvelle formule de détermination des moments d'ordre supérieur introduite dans ce chapitre ouvre des perspectives encourageantes. Lors de la détermination des moments d'ordre supérieur à partir des matrices de co-moments des titres, de nombreux termes sont encore redondants. Notre nouvelle formulation permettrait de ne conserver que la partie utile de ces matrices et réduire à nouveau leurs tailles, et donc le temps d'exécution nécessaire à la détermination des moments du portefeuille.

De plus, puisque les coefficients des fonctions d'utilité mobilisées (définissant la préférence pour les moments) ne dépendent que du premier moment du rendement ; davantage de recherches sur les fonctions d'utilité en générale sont nécessaires et en particulier sur les déformations subjectives des densités objectives effectuées par les investisseurs (*Cf.* Kahneman et Tversky, 1979 ; Chauveau et Nalpas, 2008 ; Kliger et Levy, 2008) dans le cadre de fonctions d'utilité dépendantes des rangs (*Cf.* Cenci et Filippini, 2006 ; Polkovnichenko, 2005).

Par ailleurs, l'utilisation des *L-moments* combinée à celle des fonctions de pénurie autori-

serait la définition d'un ensemble de mesures de performance qui permettrait, entre autres, d'unifier les approches de Briec et Kerstens (2007) et de Darolles *et alii* (2008).
Une autre extension possible serait de généraliser le modèle d'évaluation des actifs Gini-*CAPM* (*Cf.* Shalit et Yitzhaki, 1989 ; Okunev, 1990 ; Benson *et alii*, 2003). L'introduction des moments linéaires pourrait également être effectuée dans un cadre cette fois conditionnel, en utilisant des processus de type *GARCH* pour la spécification des *L-covariances* (*Cf.* Jondeau et Rockinger, 2003-a, 2008 ; Jondeau *et alii*, 2007), ou encore en utilisant des modèles de quantile dynamique autorégressif (par exemple le modèle *CAViaR* de Engle et Manganelli (2004), le *QAR* de Koenker et Xiao (2006), le *CHARN* de Martins-Filho et Yao (2006), le *DAQ* de Gouriéroux et Jasiak (2008) ou encore les *CARE* et *EWQR* de Taylor (2008-a et 2008-b)).

Résumé du chapitre sur "*Outliers Detection, Correction of Financial Time-series Anomalies and Distributional Timing for Robust Efficient Higher-order Moment Asset Allocations*"

La détection et le traitement des données aberrantes sont un préalable important des études statistiques et économétriques. En effet, de manière générale, la présence de données aberrantes (en anglais, *outliers*) peut engendrer des problèmes de spécifications, des prévisions erronées et une inférence approximative. Les commentaires de Bernoulli (1777) suggèrent que l'élimination des données « discordantes » était une pratique assez commune il y a plus de deux cents ans. Les ouvrages de Barnett et Lewis (1978) et Hawkins (1980) offrent à cet égard une présentation générale des premières méthodes de détection et de traitement de ces données aberrantes.

La détection des valeurs anormales est un problème très général, les causes de la présence d'*outliers* étant en effet multiples. Il peut s'agir par exemple de pannes matérielles, d'erreurs de mesure, de manipulations frauduleuses de données, d'erreurs humaines ou plus simplement de déviations « naturelles » de la population... La détection de valeurs aberrantes, la détection d'anomalies, la détection de bruit et la détection de déviations sont autant d'appellations de techniques traitant de la détection d'erreurs au sens large.

Un intérêt particulier est porté sur ces techniques de détection de réalisations atypiques puisqu'elles rendent possible la correction de ces données avant que leur utilisation n'ait des conséquences dommageables. Il semble pourtant qu'il n'y ait pas, à notre connaissance, de définition universelle des données aberrantes. Le concept général de Grubbs (1969) tend toutefois à s'imposer :

« *An outlying observation, or "outlier", is one that appears to deviate markedly from other members of the sample in which it occurs.* »

Frank Grubbs, 1969, "Procedures for Detecting Outlying Observations in Samples", *Technometrics 11*, p.1.

Grubbs (1969) propose ainsi de considérer comme aberrante toute réalisation d'une variable qui s'éloigne « sensiblement » de sa moyenne. A partir de ce concept, différentes techniques ont été proposées pour détecter ces erreurs. Hodge et Austin (2004) proposent une synthèse très complète des différentes méthodes existantes. Ces derniers les regroupent en trois catégories : les modèles d'apprentissage non-supervisés, les modèles supervisés et les modèles paramétriques.

Les modèles non-supervisés ont l'avantage majeur de ne pas nécessiter d'hypothèse forte *a priori*. Ils s'apparentent très largement aux méthodes de regroupement (en anglais *clustering*) non-supervisés (telles que les cartes de Kohonen discutées dans le Chapitre 1). Une fois les données classifiées, les observations qui constituent des groupes homogènes sont alors considérées comme normales ; inversement, les réalisations s'éloignant sensiblement de celles des autres individus du groupe sont considérées comme aberrantes, ou plutôt « non normales ».

Les modèles d'apprentissage supervisé doivent disposer d'un ensemble d'observations préalablement labellisées. A partir de sous-échantillons pour lesquels la normalité (ou anormalité) a été établie, ces modèles apprennent à différencier le type de réalisations (normales ou aberrantes).

Enfin, le troisième type de techniques modélise explicitement la normalité, au sens strict ou non. Il s'agit généralement de spécifier la distribution des données considérées normales et d'en déduire la (non) normalité au regard de la probabilité associée à cette réalisation.

En finance, les séries de rendements d'actifs sont couramment sujettes à ce problème de valeurs aberrantes. Dès 1974, Rosenberg et Houglet insistent sur leur fréquence, et l'impact qu'ont ces valeurs sur les modèles financiers. Le traitement de ces rendements atypiques constitue ainsi une problématique incontournable. Lorsqu'il s'agit de déterminer les ren-

dements potentiellement atypiques présents dans des séries d'actifs, la troisième approche (modélisation explicite de la normalité) est généralement retenue. Dans ce cas, l'hypothèse du mouvement Brownien et le paradigme Gaussien constituent le cadre standard. Mais retenir le concept de Grubbs revient alors à considérer toute réalisation extrême comme anormale. Le terme d'anormalité contient alors une certaine ambiguïté. Il peut refléter le fait que la réalisation du rendement est non-gaussienne, ou simplement improbable. Avant de supprimer ou corriger ces valeurs extrêmes, il apparaît ainsi légitime de s'interroger sur l'essence même du concept de données aberrantes en finance. Comme anticipé par Mandelbrot (1963), de nombreuses constatations empiriques ont conduit à invalider les hypothèses sous-tendant le paradigme Markov-Gaussien. Afin d'ancrer notre raisonnement, nous citons ce même auteur :

"The risk-reducing formulas behind portfolio theory rely on a number of demanding and ultimately unfounded premises. First, they suggest that price changes are statistically independent from one another... The second presumption is that all price changes are distributed in a pattern that conforms to a standard bell curve.
Do financial data neatly conform to such assumptions ? Of course, they never do."
Benoît Mandelbrot, 1999, "A Multifractal Walk down Wall Street", *Scientific American 280*, p 71.

De par sa formulation, le concept de valeur aberrante est directement sujet au problème de l'induction dans la découverte scientifique (*Cf.* Popper, 1934) puisqu'il nécessite d'effectuer une hypothèse subjective sur la distribution des rendements. Popper reprend ainsi les discussions amorcées par Hume (1748) sur les limites du principe de l'induction et du positivisme scientifique, pour qui aucune série finie d'observations ne peut permettre d'affirmer l'universalité d'un phénomène observé. Karl Popper utilise alors la métaphore de la « théorie » des cygnes blancs falsifiée lors de la découverte de cygnes noirs en Australie en 1697. Taleb (2007) reprend cette image pour l'appliquer aux évènements de marché ; il définit un « cygne noir » en finance comme un évènement rare (dans la mesure où aucune réalisation passée ne s'apparente à cet évènement). Le krach financier du lundi noir est un exemple parfait de ce type de réalisation. Le 28 octobre 1929, l'indice Dow

Jones plongeait de plus de 12%. Une telle baisse n'était survenue depuis la création de l'indice le 26 mai 1896. Sous l'hypothèse distributionnelle de normalité des rendements de l'indice, cet évènement est définitivement hautement improbable. Faut-il alors considérer cette réalisation comme aberrante ou rejeter le postulat de la normalité des rentabilités des actifs ? Nous envisagerons cette question dans le Chapitre 3.

Il nous semble en effet nécessaire de considérer des hypothèses moins restrictives que le simple paradigme Gaussien lors de la définition du concept de données aberrantes. Ce cadre élargi doit notamment permettre de différencier les réalisations extrêmes des réalisations atypiques. En effet, comme le soulignent Borland *et alii* (2005), ces anomalies sont cruciales pour l'estimation des risques financiers.

L'estimation précise des moments des distributions d'actifs est un élément primordial lors de la réalisation de modélisations financières. En effet, de très nombreux modèles financiers se fondent sur les moments des rentabilités des actifs financiers ; les erreurs d'estimation des moments empiriques vont ainsi se répercuter directement sur les résultats des modèles. L'hypothèse de normalité des rendements permettant de décrire l'ensemble des distributions à partir des deux premiers moments, de nombreux modèles se fondent sur les espérances et les (co)variances des rendements des actifs. L'impact des rendements anormaux sur les portefeuilles efficients dans le cadre traditionnel de l'espérance-variance a ainsi été largement étudié. Best et Grauer (1992) ou encore Chopra et Ziemba (1993) par exemple, évaluent l'impact de l'erreur de l'estimation des moments sur les portefeuilles efficients. Ils concluent qu'une erreur d'estimation sur l'espérance des rendements a des conséquences plus marquées sur les portefeuilles efficients qu'une erreur sur la variance ou sur les covariances.

La constatation du caractère asymétrique et leptokurtique des distributions des rendements des actifs ont incité les chercheurs académiques à prendre explicitement en compte les moments d'ordre supérieur (*Cf.* Lai, 1991 ; Chunhachinda *et alii*, 1997 ; Wang et Xia, 2002 ; Chang *et alii*, 2003 ; Sun et Yan, 2003 ; Jondeau et Rockinger, 2003-b ; Jurczenko

et Maillet, 2006-a ; Jurczenko *et alii*, 2006) pour déterminer l'ensemble des portefeuilles efficients. Or ces moments sont extrêmement sensibles à ces rendements atypiques (*Cf.* par exemple, White *et alii*, 2008 ; Jondeau et Rockinger, 2009). En raison de cette sensibilité extrême des mesures d'asymétrie et de *kurtosis* aux rendements atypiques, certains auteurs (Kim et White, 2004) estiment que les moments d'ordre supérieur sont virtuellement déterminés par des erreurs de données et que la prise en compte de ces derniers est vaine.

Nous avons proposé un modèle d'allocation d'actifs intégrant les quatre premiers moments (traditionnels, *Cf.* Annexe 4, et robustes, *Cf.* Chapitre 2) des distributions des rendements des portefeuilles. Nous poursuivons donc cette analyse dans ce troisième chapitre en étudiant l'impact des données aberrantes présentes dans les séries temporelles des actifs financiers sur les modèles d'allocation d'actifs dans le cadre moyenne-variance-*skewness*-*kurtosis*. Plus précisément, notre objectif consiste à établir une différenciation entre les réalisations extrêmes et les rendements aberrants. En cela, notre contribution s'insère dans un champ de la littérature relativement récent, qui consiste à identifier parmi les événements extrêmes ceux qui le sont réellement (*Cf.* Mittnik *et alii*, 2000 ; Huisman *et alii*, 2001 ; Johansen et Sornette, 2001 ; Gonzalo et Olmo, 2004). Nous supposons qu'une réalisation extrême, lorsqu'elle survient, a un impact sur le marché dans son ensemble.

Parmi les articles traitant du problème des données aberrantes dans la détermination des portefeuilles efficients (*Cf.* Michaud, 1998, par exemple), deux types d'approches sont généralement proposés.

La première approche consiste à substituer aux moments traditionnels des mesures dites « robustes » (*Cf.* Ledoit et Wolf, 2003 ; Martellini et Ziemann, 2007). Celles-ci peuvent être obtenues au moyen de techniques de ré-échantillonnages (*Cf.* Michaud, 1989 ; Kosowski *et alii*, 2007), ou directement en considérant des statistiques robustes (*Cf.* Yitzhaki, 2003 ; et Kim et White, 2004). Suivant cette approche, nous avons proposé dans le Chapitre 2 un modèle d'allocation d'actifs intégrant les quatre premiers moments des distributions

des rendements des actifs dans un cadre robuste. Les *L-moments* (*Cf.* Hosking, 1990) constituent une alternative permettant de palier les limites des moments traditionnels. Ce modèle ayant été largement présenté au cours du Chapitre 2, nous nous intéressons plus en détail dans la suite à la seconde approche.

La seconde voie consiste alors à retraiter en amont la base de données. Il s'agit de la filtrer, en détectant les valeurs dites « anormales » et en les corrigeant, avant d'utiliser de nouvelles séries re-traitées lors de la phase d'estimation. Depuis la typologie *princeps* de Fox (1972) élargie par Chen et Liu (1993), quatre types d'*outliers* sont généralement distingués :

- un point aberrant ponctuel qui affecte uniquement une observation de la série, et non ses valeurs futures (*Additive Outlier*),
- un point aberrant progressif (*Innovative Outlier*) qui affecte temporairement le processus à la manière des innovations considérées classiquement en analyse des séries temporelles,
- un changement transitoire (*Transitory Change*) qui affecte le niveau de la série à une certaine date mais dont l'influence décroît ensuite de manière exponentielle,
- un changement de niveau (*Level Shift*) qui affecte toutes les observations d'un impact constant à partir d'une certaine date.

Le premier type de réalisations extrêmes est généralement considéré comme émanant d'un changement exogène, alors que le second est supposé être lié à un choc endogène. Par ailleurs, les deux premières catégories concernent généralement des observations atypiques, alors que les deux dernières catégories sont associées à des changements structurels éphémères ou permanents.
Franses et Ghijsels (1999) adoptent une approche originale pour déterminer les rendements atypiques de type *Additive Outliers*. Plutôt que de considérer directement la distribution des rendements, ils proposent de modéliser explicitement le processus de volatilité puis

de définir les réalisations aberrantes à partir de l'impact que celles-ci ont sur la volatilité. Ils retiennent la modélisation de la volatilité par un processus stationnaire autorégressif avec hétéroscédasticité conditionnelle (*GARCH* de l'anglais *Generalized AutoRegressive Conditional Heteroskedasticity*) proposée par Engle (1982) et Bollerslev (1986) qui permet de rendre compte du phénomène de regroupements temporels des grandes volatilités et modéliser les effets des variations des volatilités conditionnelles. Dans ce cadre, Baillie et Bollerslev (1989) observent que la volatilité résiduelle issue de ces modèles *GARCH* est fréquemment caractérisée par un coefficient d'aplatissement supérieur à celui d'une distribution normale. Balke et Fomby (1994) expliquent le caractère leptokurtique de ce résidu par la présence de rendements aberrants au sein de la série.
Ainsi, Franses et Ghijsels (1999) proposent un modèle de détection et de correction des rendements atypiques fondé sur une modélisation de la volatilité par une approche de type *GARCH* et dont les réalisations atypiques sont déterminées à partir du résidu de la modélisation de la volatilité. Cette approche a l'avantage notable de rendre compte des changements de niveau (*Level Shift* et *Transitory Change*) caractéristiques de la volatilité. Ainsi les périodes de fortes turbulences de marché, caractérisées par de fortes volatilités, sont explicitement intégrées dans leur modélisation. Les réalisations extrêmes ne seront ainsi pas nécessairement considérées comme aberrantes. Ce premier modèle permet en particulier la détection et la correction des *outliers* additifs. Charles et Darné (2005) ont récemment proposé une extension de ce modèle permettant de détecter aussi les réalisations atypiques dites « innovantes » (*Innovative Outliers*).

Suivant l'approche de Charles et Darné (2005), nous proposons de substituer au modèle *GARCH* utilisé une classe plus générale de modèles : les réseaux de neurones *GARCH* (noté *ANN-GARCH*, de l'anglais *Artificial Neural Network GARCH*). Introduits par Donaldson et Kamstra (1997), les *ANN-GARCH* se décomposent en un modèle *GARCH* traditionnel auquel s'ajoute un réseau de neurones de type « perceptron multicouche ». L'objectif de la composante neuronale est de modéliser la partie non-linéaire prévisible présente dans le résidu du modèle *GARCH*. En finance, les *ANN-GARCH* sont utilisés

pour modéliser et prévoir la volatilité des actifs. Les différentes études comparant les performances des *GARCH* aux *ANN-GARCH* concluent que l'apport des réseaux de neurones permet le plus souvent une modélisation plus réaliste de la volatilité, tout en réduisant significativement l'erreur de modélisation ou de prévision (*Cf.* Dounias et Thomaidis, 2008). Par ailleurs, l'introduction de cette composante non-paramétrique permet de réduire les risques de modèle et ainsi d'améliorer sensiblement la détection des rendements aberrants.

Pour illustrer et tester notre modèle de détection et correction des points aberrants, nous considérons une base de données quotidiennes comprenant le prix des principaux actifs composant l'indice CAC40. L'étude couvre la période du 1^{er} janvier 1996 au 21 janvier 2009. Afin de disposer d'une base complète, nous supprimons les 11 titres dont l'inclusion au sein de l'indice a été réalisée après le premier janvier 1996. Cette base de prix, récente, a la particularité d'inclure plusieurs crises (1997, 1998, 2001 et 2008) ainsi qu'un cycle économique complet. Elle semble ainsi particulièrement adaptée à l'étude des valeurs extrêmes. En outre, les tests réalisés sur les séries de rendements conduisent clairement à rejeter l'hypothèse de normalité et ils valident sans ambiguïté le caractère autorégressif des volatilités conditionnelles pour l'ensemble des séries étudiées.

Nous comparons, dans un premier temps, les différences induites par la substitution d'un modèle *ANN-GARCH* à un modèle *GARCH*. En dépit d'une réduction de l'erreur résiduelle du modèle de volatilité de l'ordre de 11%, l'utilisation des *ANN-GARCH* ne semble pas avoir d'impact significatif sur la détection des rendements atypiques potentiels. Parmi les 178 réalisations atypiques révélées par le modèle de Charles et Darné (2005), 174 le sont également par notre modèle. Nous étudions les différences constatées entre les rendements reconstruits. Le choix du modèle (*GARCH* ou *ANN-GARCH*) ne semble pas avoir de conséquences majeures sur ces séries reconstruites. Seules les séries dont la distribution est particulièrement leptokurtique ont des rendements reconstruits significativement différents suivant la méthode utilisée. Nous choisissons néanmoins de retenir la détection des *outliers* par *ANN-GARCH*.

Une première observation étonnante provient du fait que parmi les principaux rendements

aberrants détectés, quasiment aucun ne survient parmi les crises majeures identifiées. Nous cherchons alors une justification économique aux 174 *outliers* détectés et une analyse qualitative est conduite. A partir d'archives de dépêches boursières et financières, nous résumons l'information ayant engendré les rendements extrêmes. Dans une très large majorité des cas, nous établissons que ces rendements atypiques proviennent d'informations concernant les sociétés, et non une simple conséquence de turbulences généralisées de marché. Une analyse quantitative complémentaire est également conduite. Nous nous proposons d'utiliser un indicateur de chocs de marché (*IMS*, pour *Index of Market Shocks* en anglais, *Cf.* Maillet et Michel, 2003) pour comparer les chocs observés pour un titre spécifique aux turbulences générales de marché. Cette seconde analyse nous permet de valider le caractère idiosyncratique des rendements atypiques détectés.

Nous abordons ensuite l'étude de l'impact de ces données aberrantes sur les modèles d'allocation prenant en compte les moments d'ordre supérieur. Nous orientons notre analyse sur la composition et les quatre premiers moments d'un ensemble de portefeuilles efficients qui soient réalistes aux regards des préférences des investisseurs. En effet, il est peu probable qu'un portefeuille d'asymétrie maximale ou de coefficient d'aplatissement minimal soit retenu par un agent rationnel, indépendamment de toute considération sur les autres moments. Nous choisissons ainsi de restreindre notre analyse aux portefeuilles dont les volatilités ne sont pas statistiquement différentes de celle du portefeuille de volatilité minimale. Dans cet ensemble restreint, nous cherchons les portefeuilles (locaux) d'espérance et d'asymétrie maximale ainsi que ceux de volatilité et de *kurtosis* minimum. A partir des bases brutes ou retraitées par la méthode de correction-détection issue d'un modèle *ANN-GARCH*, nous cherchons les portefeuilles qui possèdent des caractéristiques remarquables (efficients dans l'espace moyenne-variance-*skewness*-*kurtosis*) en considérant les moments traditionnels ou robustes. Nous retenons une fenêtre glissante d'un an pour réaliser ensuite une simulation dynamique de quatre portefeuilles cibles en adoptant un re-balancement mensuel. L'étude des séries de rendements des portefeuilles dynamiques nous permet enfin d'évaluer l'apport du débruitage et des statistiques robustes.

Au regard de nos tests, les propriétés des portefeuilles obtenus sont significativement améliorées lorsque l'on considère la base de données corrigée des rendements aberrants. A l'exception du portefeuille de volatilité minimale pour lequel aucune des deux techniques envisagées ne conduit à une amélioration quelconque. En ce qui concerne les *L-moments*, nos conclusions sont plus mitigées. Ils sont certes moins sensibles aux points aberrants, et donc plus stables que les moments traditionnels ; mais cette stabilité semble induire une certaine inertie qui pénalise les portefeuilles dynamiques obtenus.

Nous envisageons ainsi de poursuivre cet axe de recherche en exploitant directement la variabilité des moments empiriques. Plusieurs voies sont dès lors possibles.

La première amélioration potentielle consistera à utiliser un réseau de neurones dit « récurrent », afin de mieux s'adapter à la dynamique des moments estimés. Le modèle non-linéaire de prévision de la volatilité proposé par Miazhynskaia *et alii* (2006) pourra ainsi être repris et développé pour améliorer les prévisions de notre perceptron multi-couches.

L'exploitation des variations des moments conditionnels des rendements des actifs nous incite naturellement à examiner plus avant les récents travaux de Jondeau et Rockinger (2008). Ces derniers généralisent le concept « d'anticipation de la volatilité » de marché (« *volatility timing* », en anglais) en introduisant l'idée d'une anticipation des quatre premiers moments des distributions des rentabilités des actifs (« *distributional timing* », en anglais). La formalisation de telles anticipations rend nécessaire de disposer d'estimations conditionnelles de volatilité, mais également des cœfficients d'asymétrie et d'aplatissement des queues de distributions. La prise en compte de ces anticipations nous permettra ainsi d'améliorer notablement notre exercice d'allocation d'actifs dynamique en intégrant directement les anticipations des agents.

Une dernière piste d'amélioration consistera à étudier l'impact des différents paramètres considérés dans notre étude. La taille de la fenêtre glissante retenue dans notre exercice d'allocation dynamique pourra ainsi sensiblement influer sur nos résultats. Nous envisageons également une étude approfondie sur les conséquences du choix de la valeur critique permettant de différencier les valeurs extrêmes des valeurs aberrantes.

Enfin, l'étude des valeurs extrêmes pourra enfin être complétée par la considération de

lois de distribution plus adaptées. Nous pensons notamment utiliser une distribution de Pareto généralisée pour évaluer les probabilités d'occurrence des rendements extrêmes ou aberrants détectés.

Il sera finalement intéressant de comparer les portefeuilles dynamiques obtenus après une reconstruction de type *ANN-GARCH* avec des méthodes plus classiques de reconstruction des valeurs manquantes (telles que celles présentées au Chapitre 1 par exemple).

Résumé du chapitre sur “Carte de style et facteurs de risque”

La référence à un « style » de gestion est aujourd'hui une pratique largement répandue dans l'industrie de la gestion collective. Les résultats d'études académiques (*Cf.* Basu, 1983 ; Fama et French, 1992) ont fondé les bases de stratégies d'investissement incontournables pour l'industrie. Ils ont, en outre, profondément modifié les analyses d'évaluation et d'attribution de performance. Les styles traditionnels de gestion reposent essentiellement sur une caractérisation des stratégies financières systématiques mises en place par les gérants. Lorsque l'on considère des portefeuilles investis en actions, ces stratégies consistent à privilégier la sélection de sociétés disposant de certaines propriétés (faible ratio bénéfice-cours, taille de la capitalisation, croissance du chiffre d'affaires, *etc.*). Les gérants sont régulièrement amenés à communiquer sur le style de leurs fonds. Ainsi, la connaissance du style offre aux investisseurs une meilleure compréhension du comportement des fonds et des risques associés. En effet, les performances des styles semblent être liées au déroulement du cycle d'activité (*Cf.* Fama et French, 1989 ; Gertler et Gilchrist, 1994 ; Liew et Vassalou, 2000 ; Barberis et Shleifer, 2003). En dépit de cet engouement pour la caractérisation des styles de gestion, de nombreux problèmes techniques se posent lors de la détermination du style d'investissement des fonds. Tout d'abord, il n'existe pas de définition rigoureuse des styles (chaque fournisseur d'indice retient ainsi sa propre méthodologie pour définir ses indicateurs de style). Ensuite, les procédures d'identification se fondent sur différentes hypothèses fortes, difficilement vérifiables en pratique (indépendance des facteurs, stabilité des styles...). Enfin, peu de modèles autorisent la prise en compte de la dynamique des styles retenus par le gérant (*Cf.* Chan *et alii*, 2002 ; Annaert et van Campenhout, 2007 ; Monarcha, 2008), alors que de nombreuses études concluent à des changements temporels du style des gérants (*Cf.* Swinkels et van der Sluis, 2001 ; Kim *et alii*, 2005) ayant des répercutions directes sur les performances des fonds (*Cf.* Cremers et Petajisto, 2007 ; Huang *et alii*, 2008).

Dans son article fondateur, Sharpe (1988) propose d'introduire la notion de style pour caractériser les orientations de gestion. Il suppose que les différences de comportements des gérants se traduisent directement par les stratégies retenues et qu'elles se répercutent sur leurs performances. Les performances sont ainsi considérées comme une réalisation des facteurs de styles. Dans ce cadre, ces facteurs doivent ainsi pouvoir être extraits de l'historique des rendements des fonds. Sharpe (1992) propose alors une approche statistique pour déterminer les styles prépondérants des portefeuilles. Ce modèle reste aujourd'hui encore la référence en la matière dans l'industrie de la gestion collective.

La caractérisation – ou discrimination – du style d'un fonds est néanmoins une pratique bien antérieure aux travaux de Sharpe. En effet, les définitions des catégories de fonds d'investissement étant suffisamment vagues pour autoriser un large panel de stratégies d'investissement ; la nécessité de caractériser plus en détail les stratégies adoptées par les gestionnaires s'est rapidement imposée.

Les modèles d'analyse de style ont vocation à décrire au mieux les spécificités des actifs financiers. Outre l'analyse détaillée de la stratégie d'un gérant, les modèles d'analyse de style rendent possible la détermination d'indices de gestion *ad hoc*. Ces indices personnalisés permettent aux gestionnaires de définir plus précisément leurs objectifs et se démarquer des autres fonds de leur univers de gestion. Ils sont alors en mesure de justifier des différences de performances avec l'indice général de leur catégorie. Pour un investisseur, ces indices sur mesure autorisent une analyse plus fine de l'aptitude des gérants à générer de la performance dans un cadre de gestion bien défini.

En outre, dans la mesure où le style annoncé va être déterminant lors de l'attribution de performance, les gérants peuvent être incités à communiquer sur l'adoption d'un style qui leur est favorable. Certaines études empiriques montrent ainsi l'existence d'écarts entre les styles de gestion déclarés et ceux effectivement appliqués par le gérant étudié (*Cf.* Brown et Goetzmann, 1997 ; diBartolomeo et Witkowski, 1997 ; Kim *et alii*, 2005).

L'utilité des modèles d'analyse de style est accrue lorsque l'on considère les fonds alternatifs. Les objectifs de gestion sont généralement la performance absolue avec une exposition directe limitée aux marchés d'actifs traditionnels. Il n'y a pas d'indices de référence

consensuels permettant de refléter le positionnement rendement-risque visé par le fonds. Les gestionnaires alternatifs sont, de plus, peu enclins à dévoiler leurs stratégies. Il apparaît alors nécessaire de disposer d'un modèle d'analyse de style permettant la détermination d'indices de gestion adéquats. Finalement, l'identification des facteurs de style permet aux investisseurs d'acquérir une meilleure connaissance des risques effectivement supportés (*Cf.* Fung et Hsieh, 1997 ; Lhabitant, 2002). A travers une meilleure évaluation de la performance ajustée au risque des fonds, les méthodes d'analyse de style peuvent également offrir aux investisseurs de nouvelles opportunités de gestion active. Ammann et Verhofen (2006), par exemple, montrent comment l'analyse de style permet aux investisseurs de définir des allocations tactiques en fonction des configurations de marché. Enfin, dans le cadre de la gestion passive, de tels modèles peuvent être mobilisés de façon à construire des portefeuilles diversifiés en termes de facteurs spécifiques (*Cf.* de Roon *et alii*, 2004).

Suivant l'approche originale de Rousset et Maillet (2003), nous proposons dans ce chapitre d'utiliser les cartes de Kohonen (*Cf.* Kohonen, 1995 et le Chapitre 1) pour définir un modèle d'analyse de style. Après une présentation des faits stylisés empiriques à l'origine de la considération des styles caractéristiques des marchés d'actions, nous présentons les modèles d'analyse de style les plus répandus. Nous évoquons ensuite les limites auxquelles font face ces derniers avant de présenter notre méthode de détermination des styles par les cartes de Kohonen. Nous montrons en particulier comment celles-ci permettent de surmonter certaines des principales difficultés rencontrées par les modèles usuels. Nous proposons alors une nouvelle évaluation des biais des styles de gestion annoncés. Au moyen de projections successives des fonds sur la carte, nous étudions enfin la dynamique des styles retenus par les gérants.

Les modèles d'analyse de fonds ont évolué en lien étroit avec les modèles d'évaluation des actifs. Les premières tentatives de caractérisations des fonds s'attachaient principalement à créer des groupes homogènes en termes d'exposition aux risques de marché. Ces modèles regroupaient les fonds en fonction de leur risque systématique issu du Modèle

d'Evaluation Des Actifs Financiers (*CAPM* en anglais ; *Cf.* Sharpe, 1964) pour regrouper et tenter d'expliquer les rendements des différents groupes de fonds. Avec l'émergence de l'« *Asset Pricing Theory* » (*APT*, *Cf.* Ross, 1976), apparaît une nouvelle génération de modèles prenant en compte plusieurs facteurs. Une très large majorité des modèles de style actuels, le modèle de Sharpe notamment, s'inscrit toujours dans ce cadre multi-factoriel. Ces derniers se différencient essentiellement par leur choix et l'identification des facteurs. En amont de chacun de ces modèles, l'élément déterminant réside ainsi dans la définition des facteurs de style. Le choix de ces derniers dépend directement du type d'actifs que l'on souhaite analyser.

Lorsque l'on considère des marchés d'actions, la définition des facteurs représentatifs a été largement influencée par l'observation d'anomalies de marché. Une anomalie peut être ici définie comme un phénomène constaté empiriquement sur les marchés, mais inexpliqué par les modèles d'évaluation d'actifs.

Par exemple, les entreprises de petite capitalisation démontrant empiriquement une capacité à générer des rendements supérieurs à celles de grosse capitalisation, de nombreuses études ont conclu à l'existence d'une prime « *Small Cap* » (*Cf.* Lakonishok et Shapiro, 1986), c'est-à-dire d'un biais dans les rendements des actions en fonction de la taille de la capitalisation.

L'étude des liens entre les ratios comptables des entreprises et l'évolution de leurs cours de bourse a également donné lieu à l'introduction de nouveaux styles. Lakonishok *et alii* (1994) observent par exemple que le ratio « valeur comptable sur cours » (*Book-to-Market*) est déterminant lors de l'estimation du rendement attendu d'un titre. Fama et French (1998) constatent en particulier que les entreprises dont le ratio « valeur comptable sur cours » est faible, disposent d'un potentiel de gain supérieur à celui du marché. Ils concluent ainsi à l'existence d'une prime « *Value* ».

La prise en compte de ces anomalies a donnée lieu à la définition de quatre facteurs explicatifs pour les marchés actions : les styles *Growth*, *Value*, *Large* et *Small*. Ces quatre styles constituent aujourd'hui les facteurs principaux auxquels se réfèrent les gérants de

fonds investis en actions.

Si les définitions des styles *Large* et *Small* sont immédiates (à partir des niveaux de capitalisations des titres détenus), celles des styles *Growth* et *Value* sont plus incertaines et toujours discutées. La première distinction entre les styles *Growth* et *Value* trouve son origine dans les travaux de Graham (1949). Une entreprise est qualifiée de *Value* lorsque son prix de marché est assorti d'une décote par rapport à sa valeur intrinsèque. Cette dernière est définie comme étant égale à la différence entre la valeur des actifs circulants de la firme, diminuée de celle de l'ensemble de ses dettes. Cet indicateur vise à mesurer l'aptitude de la firme à continuer de distribuer un flux régulier de dividendes. Il s'agit souvent d'entreprises opérant dans des secteurs traditionnels avec une croissance de leurs activités (revenus et profits) lente et régulière. Les valeurs *Growth* caractérisent les titres bénéficiant d'un taux de croissance historique et prévisionnel des bénéfices nettement supérieur à celui du marché. Elles disposent généralement de faibles ratios *Book-to-Market* et de taux de distribution de dividendes également faibles. Il s'agit typiquement de sociétés innovantes et avant-gardistes dans des secteurs modernes et dynamiques. Empiriquement, Lakonishok *et alii* (1994) observent que les firmes à fort *Book-to-Market* ont de faibles bénéfices, mais procurent des rendements moyens supérieurs à ceux constatés en général sur le marché (tous styles confondus). Inversement, les entreprises ayant un *Book-to-Market* faible, disposent de bénéfices élevés, mais de rendements faibles. L'ambigüité concernant ces styles est d'abord structurelle. En effet, en fonction des cycles économiques et des évolutions spécifiques des entreprises (*Cf.* Fama et French, 2007), le style d'un titre peut évoluer. Il résulte de ces évolutions des dérives de style dites « passives » pour les portefeuilles comprenant ces actions. Par ailleurs, au gré des cycles de marché, les facteurs de style vont offrir des perspectives de gain variables. Sur le long terme, le style *Value* semble délivrer les meilleures performances, néanmoins à plus court terme, il peut évidemment être dominé par le style *Growth*. Les gérants ayant adoptés et annoncés un style particulier sont ainsi largement incités à s'en écarter lorsque les anticipations sur leurs styles ne sont pas bonnes. On parle dans ce cas de dérives « actives ». Enfin, la complexité des méthodes d'attribution de style peut également être la source de biais lors des auto-déclarations

des styles des gérants (*Cf.* diBartolomeo et Witkowski, 1997). Ces incertitudes concernant la définition de styles « purs » ont motivé l'introduction de nuances de styles. Ainsi les fonds dont la caractéristique *Growth* est particulièrement marquée seront considérés comme « *Aggressive Growth* ». De même, le style « *Deep Value* » est attribué aux fonds intégrant des titres particulièrement sous évalués. Enfin, un style composite regroupe les fonds sans style particulier ou combinant de façon équilibrée différents styles (« *Blend* », ou *GARP* : *Growth-at-Reasonable-Price* ; *Cf.* Ainsworth *et alii*, 2008).

A l'échelle d'un portefeuille, deux grandes approches permettent la détermination des styles. La première consiste à analyser les différents titres composant le portefeuille (*Holding-based Style Analysis,* ou *HBSA* en abrégé, *Cf.* Daniel *et alii*, 1997). Dans un premier temps, les caractéristiques de chacun des actifs du portefeuille sont analysées. Pour des actions, par exemple, les ratios comptables (passés, actuels et futurs estimés) vont permettre d'en déduire les styles. Dans un second temps, les styles des titres détenus sont agrégés autorisant ainsi la détermination du style prépondérant du portefeuille analysé. La quantité d'information requise rend complexe la mise en place effective d'une telle analyse : la composition des portefeuilles n'est généralement pas divulguée (ou avec un certain retard et un problème d'asynchronisme avéré) ; les ratios nécessaires (en particulier les consensus des prévisions des analystes) sont difficilement accessibles (*Cf.* Otten et Bams, 2001). Ainsi les méthodes du type *HBSA* restent principalement utilisées par les fournisseurs d'indices ou les gestionnaires eux-mêmes et sont difficilement applicables par les autres investisseurs.

La seconde approche, introduite par Sharpe (1988 et 1992), permet de mesurer la sensibilité du portefeuille à différents facteurs de risque. Cette approche, communément appelée *Return-based Style Analysis* (*RBSA*) adopte un fondement statistique pour évaluer le style des fonds. Elle consiste en une régression contrainte des rendements du fonds sur ceux des facteurs de risque. Simple à mettre en œuvre, l'approche *RBSA* permet une interprétation directe des styles des fonds (dans la mesure où les facteurs retenus sont de vrais actifs). Du fait de leur simplicité d'implémentation, ce type de modèles s'est ainsi rapidement

imposé. Dans son article fondateur, Sharpe utilise l'analyse de style pour caractériser des fonds d'investissements globaux. Les styles retenus correspondent à des indices des grandes classes d'actifs : actions, obligations, bonds du Trésor. Avec l'émergence des styles caractéristiques *Growth*, *Value*, *Large* et *Small*, les grands fournisseurs d'indices ont ensuite développé des méthodologies (*Cf. MSCI*, *Stoxx*, *Lipper*, par exemple) permettant de décomposer les indices généraux en sous-indices de style. La commercialisation de ces indices a rendu possible l'évaluation simplifiée des sensibilités des fonds d'actions à ces facteurs. Enfin, des résultats récents tendent à montrer qu'une approche de type *HBSA* n'induit pas nécessairement de meilleurs résultats que l'approche *RBSA* (*Cf.* ter Horst *et alii*, 2004).

Les concepts de facteurs de style présentés, nous proposons de rappeler les principales méthodologies de type *RBSA*. Cochrane (2001), partant d'un modèle général d'évaluation des actifs, montre que la plupart des modèles classiques d'évaluation ne sont que des cas particuliers de celui qu'il considère. Son modèle permet de réconcilier les deux grandes approches en la matière : les modèles absolus (raisonnement à l'équilibre) et les modèles relatifs (raisonnement par arbitrage). Les modèles absolus sont définis à partir d'une économie à l'équilibre. Ils sont adaptés à l'étude des interactions entre les actifs financiers et l'économie réelle. L'approche relative consiste à déterminer la valeur des actifs à partir de prix d'autres actifs, sans chercher l'origine de ces derniers. Reprenant le cadre unifié de Cochrane (2001), nous présentons les modèles d'analyse de style les plus répandus. Cette formulation permet notamment de mettre en avant le fait que les modèles de style se différencient, avant tout, par le choix des facteurs retenus. Nous évoquons, ensuite, les critiques majeures auxquelles font face ces modèles (*Cf.* Corielli et Meucci, 2004). Les approches multifactorielles, notamment celle proposée par Sharpe (1988), imposent que les facteurs de style respectent trois conditions : être mutuellement exclusifs (les indices choisis ne doivent pas inclure un même titre), tout en étant exhaustifs (chaque indice devant être le plus large et le plus représentatif possible), et enfin linéairement indépendants (*Cf.* Sharpe, 1992 ; Lobosco et diBartolomeo, 1997 ; de Roon *et alii*, 2004). Or, une simple

analyse des facteurs traditionnellement utilisés met en évidence de fortes colinéarités entre ces indices. De plus, les contraintes portant sur les bêtas des régressions induisent un biais d'estimation (*Cf.* de Roon *et alii*, 2004). Ainsi, toutes les régressions mises en œuvre dans le cadre de l'analyse de style classique risquent d'être fallacieuses.

Outre ces problèmes d'ordre statistique, le choix du fournisseur d'indices de style a un impact significatif sur la détermination du style du fonds (*Cf.* Gallo et Lockwood, 1997). Ces divergences illustrent l'absence de définitions consensuelles des styles *Growth* et *Value*, chaque fournisseur adoptant sa propre méthodologie.

La nécessité de choisir *a priori* les indices de style caractéristiques du fonds constitue une autre limite majeure. Ben Dor *et alii* (2003) démontrent l'importance du choix de facteurs appropriés. Ils montrent en particulier comment l'omission d'un facteur significatif peut conduire à des conclusions erronées. Ainsi, une connaissance des fonds analysés est nécessaire afin de choisir les facteurs caractéristiques.

Enfin, le cadre d'analyse statique constitue une limite du modèle. En effet, les styles adoptés par les gérants sont susceptibles d'évoluer en fonction des conditions de marché. Cette dérive peut être volontaire (choix tactique du gérant, *Cf.* Annaert et van Campenhout, 2007 ; Basak *et alii*, 2007 ; Ainsworth *et alii*, 2008) ou simplement subie (*Cf.* Fama et French, 2007) en raison de l'évolution des cycles économiques et des caractéristiques intrinsèques des actifs détenus. Par exemple, les valeurs technologiques pouvaient être considérées comme *Growth* au plus haut de la bulle « Internet ». Après l'éclatement de cette bulle, ces mêmes valeurs peuvent être considérées comme des titres *Value*. Ainsi, pour un fonds géré passivement, supposer que les sensibilités des fonds aux facteurs sont constantes reste acceptable. Dès lors que nous considérons des fonds retenant des stratégies actives d'investissement, le modèle traditionnel de Sharpe (1988) ne permet pas la détermination du style de ces portefeuilles (*Cf.* Corielli et Meucci, 2004). Néanmoins, des approches conditionnelles existent : Ferson et Schadt (1996) concluent par exemple à des changements de styles d'investissement en fonction des anticipations économiques. Ces changements ont depuis été largement modélisés (*Cf.* Kim *et alii*, 2005 ; Swinkels et van der Sluis, 2006). Il apparaît ainsi essentiel de prendre en compte les dynamiques des styles

d'investissement (*Cf.* Chan *et alii*, 2002 ; Annaert et van Campenhout, 2007 ; Monarcha, 2008). Concernant les changements de style (*style drift* en anglais), il semble nécessaire de plus de différencier les dérives passives (subies par le gérant) des réorientations (*tilt* en anglais) choisies par les gestionnaires. En effet, des travaux récents (*Cf.* Ainsworth *et alii*, 2008) réévaluent les biais d'auto-déclaration en différenciant ces deux types de divergence.

Brown et Goetzmann (1997) proposent une approche originale de la caractérisation des styles de gestion. Exploitant le lien fort entre les méthodes d'analyse factorielle et les algorithmes de classification, ils proposent de réunir les fonds en groupes homogènes. A chacun de ces groupes correspond un facteur de risque (défini comme la moyenne des fonds de chaque groupe). Les facteurs étant issus de l'univers de fonds eux-mêmes, aucune hypothèse concernant ces derniers n'est requise *a priori* pour en déterminer les styles. Les changements d'orientation de gestion sont intégrés aux facteurs de risque. En contrepartie, les facteurs obtenus par ce modèle sont abstraits ; ils ne correspondent généralement pas à des actifs tangibles échangés sur les marchés financiers.

Les critiques formulées à l'encontre des modèles de style présentés nous incitent à développer une approche alternative dont les facteurs permettent de pallier les problèmes évoqués. Nous proposons, dans ce chapitre, un modèle d'analyse de style fondé sur les cartes auto-organisées (*Cf.* Kohonen, 1995 et Chapitre 1). Nous retenons une version robuste des cartes de Kohonen (*Cf.* Guinot *et alii*, 2006) qui assure la convergence de la carte vers une solution stable. Appliqué à une base de séries de rendements d'actifs, l'apprentissage des cartes robustes permet la génération simultanée de facteurs de risque spécifiques et de groupes homogènes de fonds. Les avantages de cette approche sont multiples. Tout d'abord, dans la mesure où cette classification est non-supervisée, aucune hypothèse concernant les facteurs explicatifs pressentis n'est requise. Ainsi, de façon analogue au modèle de Brown et Goetzmann (1997), l'étape cruciale de sélection des indices de style représentatifs est écartée. De plus, les facteurs de risque étant directement extraits de la base des fonds, les biais de gestion liés aux changements de style des titres (*passive style drift*, en anglais ;

Cf. Ainsworth *et alii*, 2008) est intégré dans les facteurs obtenus.

Ainsi les facteurs issus des cartes de Kohonen ne correspondent pas nécessairement à des actifs ou des indices de marché. Une fois la carte obtenue, il est ainsi nécessaire d'effectuer une correspondance entre les facteurs générés et les styles traditionnellement considérés. Une première approche consiste à projeter des indices connus (de styles ou sectoriels) sur la carte. Leurs classes d'appartenance seront ainsi représentatives des styles des indices qu'elles intègrent. En raison des fortes colinéarités des indices de styles entre eux, cette première approche n'est pas toujours satisfaisante. Nous proposons donc une approche alternative opérant à partir de classifications conditionnelles des fonds et des indices. L'hypothèse de Sharpe (1988) est alors reprise : le gestionnaire ayant retenu un style particulier doit surperformer lorsque le facteur associé domine le marché. Nous considérons, pour chaque indice standard, les périodes pour lesquelles celui-ci surperforme le marché. Des classifications, conditionnelles aux périodes retenues, sont réalisées. Les nouvelles classes conditionnelles sont alors indexées en fonction de la performance des individus représentatifs de chacune. A chaque fonds est attribué le score de la classe conditionnelle à laquelle il appartient. Le style des groupes inconditionnels est alors déterminé en considérant les scores conditionnels de l'ensemble des fonds affectés à chaque groupe.

Nous illustrons notre méthode d'analyse de style en utilisant les valeurs liquidatives hebdomadaires de la base de données de fonds d'actions américaines distribuée par *Lipper*. La base regroupe les valeurs liquidatives de 598 fonds et 86 indices représentatifs du marché. La période de classification va du 30/08/2002 au 25/05/2007. Une particularité de cette base est d'être fournie avec le style communiqué par les gérants ; il nous est donc possible d'évaluer l'importance du biais concernant le style annoncé par les gérants. La technique de *scoring* conditionnel employée nous permet d'estimer qu'environ 70% des gestionnaires adoptent des stratégies de gestion cohérentes avec les styles auto-déclarés. Dans un second temps, nous étudions la dynamique des styles. Nous nous intéressons en particulier aux fonds dont le style auto-déclaré diffère de celui estimé par

les cartes robustes. L'étude dynamique des styles de gestion de ces fonds nous permet de différencier deux grandes classes. La première regroupe des fonds anticipant avec succès les tendances de marché (les bons *market timers* en anglais). La seconde catégorie de fonds rassemble les gestionnaires adoptant essentiellement une stratégie de *momentum*. Lorsqu'une bulle spéculative apparaît sur le marché, ces gérants y participent en achetant des actifs surévalués, bien souvent jusqu'à l'éclatement de ladite bulle. La projection dynamique des fonds sur la carte, démontre des changements multiples de positionnement. Nous concluons dans ce cas à une gestion opportuniste. Ainsi environ 10% des gérants de la base anticipent convenablement les futures conditions favorables de marché, les 20% restant ne faisant que suivre les grandes tendances de marché.

L'utilisation des cartes de Kohonen pour développer des modèles d'analyse de style offre ainsi de nouvelles perspectives en ce qui concerne l'analyse de style. Nous envisageons notamment différentes améliorations ou extensions du modèle présenté.

En ce qui concerne la calibration optimale du réseau, l'étude de l'influence des paramètres régissant la carte pourra être davantage développée. Les choix des tailles (nombre de neurones) et des topologies optimales (forme cubique, sphérique) devraient permettre d'améliorer sensiblement nos analyses. Il sera également intéressant de réaliser un débruitage préalable des données (*Cf.* Bonhomme et Robin, 2008).

L'approche conditionnelle adoptée, bien que tout à fait concluante sur notre étude, pourra également être améliorée. La technique de *scoring* en particulier pourra avantageusement prendre en compte les écarts de performance entre les facteurs de style représentatifs.

La classification et l'analyse de fonds sur lesquels nous disposions d'informations qualitatives nous a permis de valider la méthodologie proposée. Les cartes de Kohonen étant parfaitement adaptées à l'étude des styles des fonds alternatifs (*Cf.* Maillet et Rousset, 2003), nous envisageons l'étude de la dynamique des stratégies d'investissement et de couverture employées par les gérants de fonds de fonds alternatifs (*Cf.* Bodson *et alii*, 2008). Le manque de transparence caractéristique de ce type de gestion rend l'analyse de style

incontournable (*Cf.* Monarcha, 2008). Il sera ainsi intéressant de créer une carte de style à partir des fonds n'adoptant qu'une unique stratégie (les fonds dit multi-stratégies et autres fonds de fonds étant, dans un premier temps, écartés). Une fois les facteurs explicités, nous serons en mesure d'étudier la dynamique et les dérives (*style drift* en anglais ; *Cf.* Ainsworth *et alii*, 2008) des stratégies retenues, en particulier celles des gérants de fonds multi-stratégies et de fonds de fonds.

Enfin, les cartes de Kohonen autorisant la classification de variables qualitatives, différentes caractéristiques opérationnelles (structuration, conflit d'intérêt potentiel et d'autres risques opérationnels ; *Cf.* Brown *et alii* 2008-a, 2008-b et 2009) pourront avantageusement être incluses aux classifications avant de définir de nouvelles mesures de performance.

Chapter 1

Hedge Fund Time-series Completion and *Scenarii* Generation for Robust Asset Allocation and Risk Measurement

1.1 Introduction

Missing values occur recurrently in asset price database. Non-quotation dates, too recent inception dates, intention not to report bad performances or mistakes of data providers are several reasons why prices may not be provided. Since most of financial models need complete and cylindrical samples, pre-processing imputation methods have to be applied. Due to hedge funds strong peculiarities (see, Ackerman, McEnally and Ravenscraft [1999]), the need of pre-treatments is intensified when considering this asset class. More precisely, because hedge funds remain free from several requirements of regulatory bodies, they are subject to various biases (see, Eling [2006]). Backfill bias directly arise from the hedge fund manager possibilities to not disclose their performances. During the beginning of a hedge funds life, the incubation period, the manager set his investment strategy and generally not disclosed its performances. When he chooses to include its fund into a hedge

funds database, he may disclose past returns (if those are favourable) and in this case, he may smooth past returns and make the fund performances appealing. Number of authors (Brown, Goetzmann and Ibboston [1999]; Fung and Hsieh [2000]; Posthuma and van der Sluis [2003], and Malkiel and Saha [2005], among others) evaluate the impact of backfill on the real returns. Even if there is no consensus on the exact impact of backfilled returns, these studies estimate that the incubation period vary from 12 to 24 months and conclude to a significant bias that overestimates the annualized return from 1.4% to 5%. These authors suggest that the track records of hedge funds corresponding to the incubation period should be removed from the dataset before any financial model computation. Consequently, the deletion of potential backfilled returns may drastically reduce the length of the available input dataset.

These facts exacerbate the need of missing values imputations and simulations methods. Methods to handle missing data and generate *scenarii* have been active areas of statistical research. Under the Gaussian hypothesis, the Expected Maximization (see Dempster, Laird and Rubin [1977]) and Monte Carlo simulations stand for the standard references. More elaborated approaches without requiring Gaussian hypothesis also exist, but these methods are generally *ad hoc* methodologies, depending on the type of time series to be filled. It is now, well established that asset returns do not follow a Gaussian distribution. Thus, such a hypothesis is no more acceptable for several asset classes and especially for hedge funds. Their returns exhibit strong peculiarities (see Agarwal and Naik [2000], and Brooks and Kat [2002]). They are also often characterized by a significant level of skewness and a high *kurtosis*.

The goal of this paper is twofold. Firstly, we provide a way to deal with missing values in order to get cylindrical and complete sample that respects initial statistical properties (first four moments and dependency structure in our case). Secondly, we propose a way to simulate realistic multivariate returns or *scenarii* ensuring, for instance, that the sample is large enough to get statistically significant model outputs. The proposed methods operate the Kohonen algorithm altogether with Empirical Orthogonal Functions and the Constrained Randomization Method. The interest of theses methods is that they do not

required any hypothesis and they are also totally data driven.

Self-Organizing Maps (SOM) (see Kohonen [1995]) algorithm constitute powerful non-linear financial classification tools. SOM aims ideally to group homogeneous individuals, highlighting a neighborhood structure between clusters in a chosen network. Furthermore, it has the kindly property to be robust in presence of missing data (see Samad and Harps [1992]). Completion algorithm derived from SOM have already been proposed (see Fessant and Midenet [2002], Wang [2003]), but since the SOM learning could lead to different neighborhood structures and since completion algorithm coming from SOM make an intensive use of this neighborhood structure, we choose to use a modified version of SOM, the Robust SOM (see Guinot, Maillet and Rousset [2006]). Initially proposed for signal processing, Empirical Orthogonal Functions (see Hanson and Lawson [1974]) correspond to a factorial decomposition method such as the principal component analysis. They also allow to handle missing values (see Boyd, Kennelly and Pistek [1994]). One of their main advantages is that they allow a continuous imputation of missing values. The main issues are that they are known to be sensitive to initialization stage. A potential issue with SOM or EOFs based completion algorithms is that they may failed to keep higher moments and the correlation structure of the original time series database. Thus, we proposed to combine Robust SOM plus EOFs and constrained randomization algorithm in order to develop a completion algorithm that respects the above statistics. We choose to compare our algorithm with the Expected Conditional Maximization (ECM) (see Meng and Rubin [1993], and Sexton and Swensen [2000]) for missing values and we show the impact of the use of an appropriate completion algorithm by performing different asset allocations after those completion algorithms.

The structure of this paper is as follow. In the second section, we present different algorithms for imputation methods (SOM, EOFs, and the Constrained Randomization Methods algorithm) and we show how to combine these algorithms to obtain our robust algorithm to fill hedge funds missing value. In the third section, we present how SOM allow us to obtain realistic returns or *scenarii*. The last section is then dedicated to empirical illustrations. We extract from the large HFR^{TM} database a set of hedge fund returns

composed with 50 funds on a 10-year period of 120 monthly values. We artificially delete different levels of data before rebuilting them. The qualities of the reconstruction are evaluated through the comparison with real returns and with those obtained *via* ECM algorithm. A comparison between our simulation process and a Gaussian Monte Carlo is also proposed. Then, we illustrate the importance of accurate reconstructions through the results provided by traditional asset allocation and risk models.

1.2 Imputation Methods

Methods to handle missing data have been an area of statistical research for many years (see Little and Rubin, [2002]). Rubin [1976] provides one of the first systematic studies on missing data. Since his seminal work on missing data, imputation methods have became an active area of statistical practice. The first proposed methods were the Listwise deletion and the Pairwise deletion. The first one consists in remove the whole time series if there exists a missing value; the second one consists in removing the whole observation (the date) of each time series. These simplistic approaches reduce drastically the size of the sample. However, more adapted solutions have thus been proposed. Even if a wrong imputation could lead to bias in rebuilt time series (*Cf.* Harel [2008]), estimation methods have supplanted these first basic methods. Many methods have been proposed, from simple mean imputation to complex simulations. Under the Gaussian hypothesis, Expectation Maximization (EM) algorithm presented in Dempster, Laird and Rubin [1977] allows us to accurately recover missing data. This method became quickly the standard. Other elaborated tools also permit missing data estimation, but they need strong assumption on the diffusion process driving the returns. Due to vast panel of strategies used in alternative industry, such a specification can difficultly be done. We present in this section, each algorithm used in our method and then the way to combine them.

1.2.1 Self-Organizing Map Algorithms

The SOM is a method that represents statistical data sets in an ordered fashion as a natural groundwork on which the distributions of the individual indicators in the set can be displayed and analyzed. It is based on the unsupervised learning principle where the training is entirely data-driven and no information about the input data is required (see Kohonen [1995]).

The SOM consists of a network, compound in n neurons, units or code vectors organized on a regular low-dimensional grid. If $I = [1, 2, ..., n]$ is the set of units, the neighborhood structure is provided by a set of neighborhood function Λ defined on I^2. The network state at time t is given by:

$$\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_n(t)], \tag{1.1}$$

where $\mathbf{m}_i(t)$ is the T-dimensional weight vector of the unit i.

For a given state $\mathbf{m}$ and input $\mathbf{x}$, the winning unit $i_w(\mathbf{x}, \mathbf{m})$ is the unit of which weight $\mathbf{m}_{i_w(\mathbf{x},\mathbf{m})}$ is the closest to the input $\mathbf{x}$.

The SOM algorithm is recursively defined by the followings steps (see Figure 1.1):

1. Draw randomly an observation $\mathbf{x}$.

2. Find the winning unit $i_w(\mathbf{x}, \mathbf{m})$ also called the Best Matching Unit (noted BMU) such that:

$$BMU = i_w[\mathbf{x}(t+1), \mathbf{m}(t)] = \underset{i \in I}{\text{Argmin}} \{\|\mathbf{x}(t+1) - \mathbf{m}_i(t)\|\}, \tag{1.2}$$

   where $\|.\|$ is the Euclidian norm.

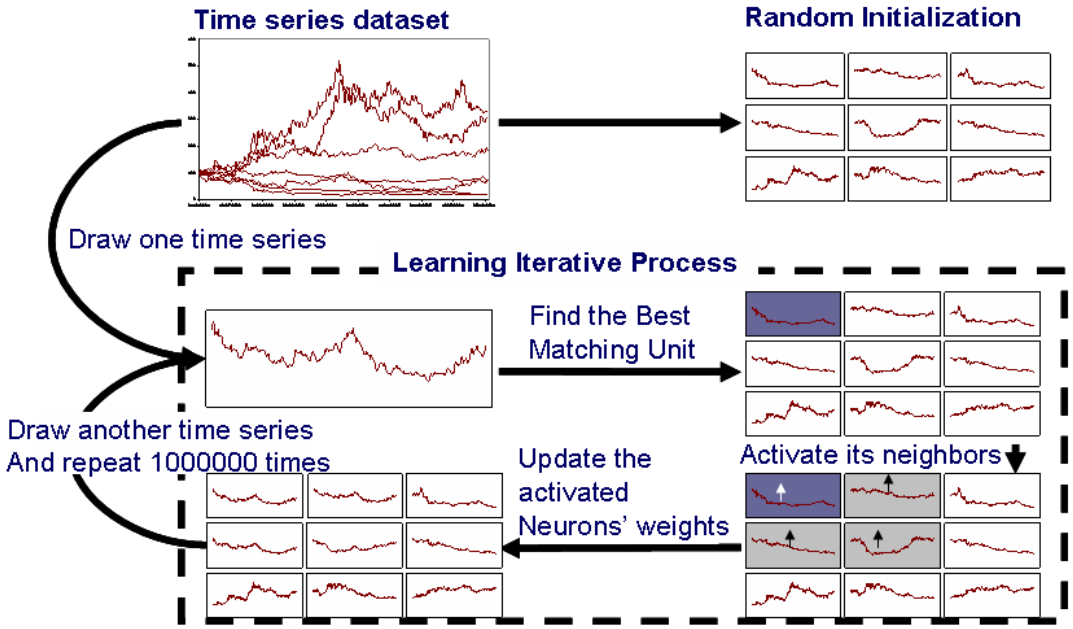
3. Once the BMU is found, the weight vectors of the SOM are updated so that the BMU and these neighbors are moved closer to the input vector. The SOM update rule is:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon_t \Lambda(BMU, i)\left[\mathbf{m}_i(t) - \mathbf{x}(t+1)\right], \forall i \in I, \tag{1.3}$$

where ε_t is the adaptation gain parameter, which is $]0,1[$-valued, generally decreasing with time.

The number of neurons taken into account during the weight update depends on the neighborhood function Λ that also decreases with time.

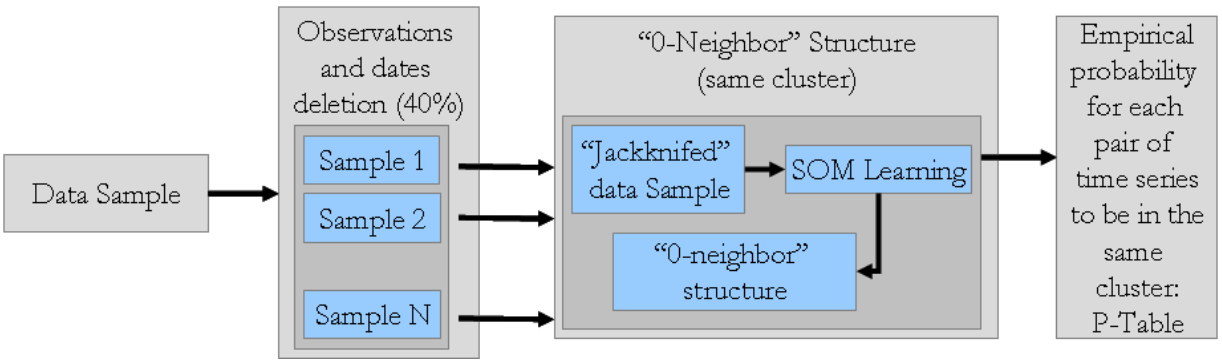
Figure 1.1. Illustration of the Kohonen Iterative Process.



When SOM are used in classification, the algorithm is applied to the complete database that is generally a sample of some unknown stationary distributions. A first concern refers to the question of the stability of the SOM solution (specifically the neighborhood organization) to changes in the sample and to contaminations by large outliers. A second concern regards the stability of the data presentation order and the initialization. To limit the dependence of the outputs to the original data sample and to the arbitrary choices within an algorithm, it is common to use a bootstrap process with a re-sampling technique (see Efron and Tibshirani [1993]). Here, this idea is applied to the SOM algorithm, when estimating an empirical probability for any pair of individuals to be neighbors in a map. This probability is estimated by the number of times the individuals have been

neighbors at ray 0 (belong to the same cluster) when running several times the same SOM algorithm using re-sampled data series (see Figure 1.2). In the following, we call **P** the matrix containing empirical probabilities for two individuals to be considered as neighbors at the end of the classification. Following Guinot, Maillet and Rousset [2006], the algorithm uses only individuals in the given re-sampled set of individuals (representing 60% or so of the original population). We generalize the previous approach by adding a drawing of observations from the original series (without replacement, around 60%) for each individuals. At the end of the first step, the remaining incomplete individuals are classified using computed distances to the code vectors. Thus, at each step, the table of empirical probabilities concerns all individuals in the original dataset, even if only a partial part of them have been used within the algorithm.

Figure 1.2. Step 1, Bootstrap Process for Building the Table **P** of the Individual's.



When the matrix **P** is built, the first step is over. In the second step (see Figure 1.3), the SOM algorithm is also executed several times, but without re-sampling. For any map M_i, we can build the table P_M, similar to the previous one, in which values are 1 for a pair of neighbors and 0 for others. Then, using the Frobenius norm, we can compute the distance between both neighborhood structures, defined respectively at the end of step 1 (re-sampling the data) and step 2 (computing several maps with the original data). The selected Robust Map, called hereafter RMap for the sake of simplicity, is the one that minimizes the distance between the two neighborhood structures as follows:

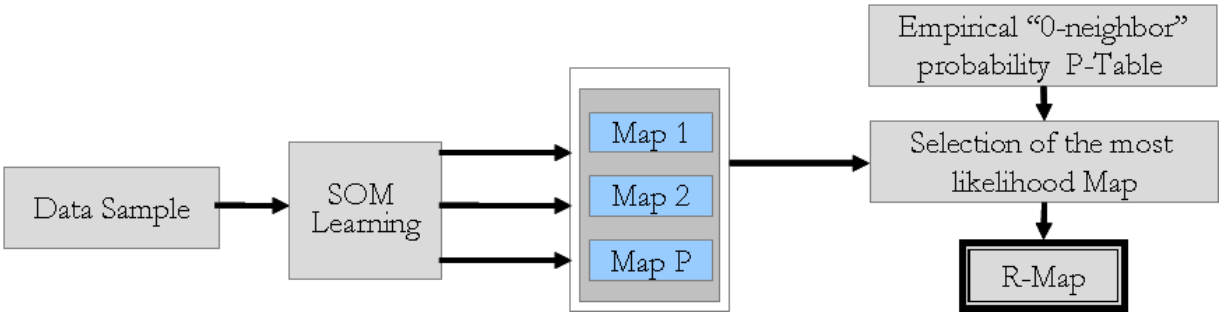
$$\mathbf{RMap} = \underset{\mathbf{M}_i, i \in I}{\text{Argmin}} \left\{ \| \mathbf{P} - P_{\mathbf{M}_i} \|_{Frob} \right\}, \tag{1.4}$$

where $\|.\|_{Frob}$ is the Frobenius norm that is:

$$\|\mathbf{A}\|_{Frob} = \frac{1}{n^2}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{[i,j]}^2}, \tag{1.5}$$

with n the dimension of the square matrix A, whose elements are $a_{[i,j]}, \forall i \in I^2$.

Figure 1.3. Step 2, RMap Selection.



The map whose neighbourhood structure is the closest to the empirical probability table P obtained at step 1, is selected.

The benefits of such a procedure are double. First the bootstrap process applied during the step one allows to minimize the effect of possible outliers present in the database. Second, the chosen RMap is the one which maximizes the likelihood of the neighborhood structure.

SOM allows for classification of data samples with multiple variables and missing values. Fessant and Midenet [2002] propose an adapted Kohonen algorithm that first clusters the data, and then replaces the missing observations. When the SOM algorithm iterates, if a vector $\mathbf{x}$ with missing value(s) is drawn, we consider the subset NM of variables which are not missing in vector $\mathbf{x}$. We define a norm on this subset (denoted $\|.\|_M$) that allows us to find the BMU (with previous notation):

$$BMU = i_w\left[\mathbf{x}\left(t+1\right), \mathbf{m}\left(t\right)\right] = \underset{i\in I}{\text{Argmin}}\left\{\left\|\mathbf{x}\left(t+1\right) - \mathbf{m}_i\left(t\right)\right\|_M\right\}, \tag{1.6}$$

with

$$\left\|\mathbf{x} - \mathbf{m}_i\right\|_M = \sum_{k\in NM}\left(\mathbf{x}_k - \mathbf{m}_{i,k}\right)^2,$$

where

$$\begin{cases} \mathbf{x}_k \text{ for } k = [1, ..., T] \text{ denotes the } k^{th} \text{ value of the chosen vector,} \\ \mathbf{m}_{i,k} \text{ for } k = [1, ..., T], \text{ for } i = [1, ..., n] \text{ is the } k^{th} \text{ value of the } i^{th} \text{ code vector,} \\ NM \text{ is the set of the net asset values } \mathbf{x}_k \text{ that are not missing.} \end{cases}$$

Once the Kohonen algorithm has converged, we get some cluster containing our time series. Fessant and Midenet [2002] proposed to fill the missing values of time-series by the corresponding values of the associated code vector.

1.2.2 Empirical Orthogonal Functions

Empirical Orthogonal Functions (EOFs, see Hanson and Lawson [1974]) gather and generalize Principal Component Analysis, Singular Spectral Analysis and they refer to Singular Value Decomposition (SVD). We choose here to use EOFs, generally applied to factorial analysis (loading factors), as a denoising tool as well as a tool for recovering missing values (Boyd, Kennelly and Pistek [1994]).

EOFs can be calculated by using standard and well-known SVD by solving the following eigenvalue problem:

$$\begin{cases} \mathbf{X}\mathbf{v} = \rho\mathbf{u} \\ \mathbf{X}^*\mathbf{u} = \rho\mathbf{v}, \end{cases} \tag{1.7}$$

where $\mathbf{X}$ is 2-dimensional data matrix, $\mathbf{u}$ and $\mathbf{v}$ are column vectors denoting horizontal and vertical eigenvectors, respectively, and ρ is the singular value (square root of the eigenvalue). The $\mathbf{X}^*$ is the adjoint of the matrix $\mathbf{X}$, which is the same as the conjugate transpose of $\mathbf{X}$. If $\mathbf{X}$ contains only real values, $\mathbf{X}^*$ is an ordinary transpose of $\mathbf{X}$.

When all eigenvectors and singular values are calculated, they are ranked in a descending order denoted as ρ_i for $i = [1, ..., n]$.

Then the original matrix $\mathbf{X}$ can be reconstructed as:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^* = \sum_{k=1}^{n} \rho_k \mathbf{u}_k \mathbf{v}_k, \tag{1.8}$$

where $\mathbf{U}$ and $\mathbf{V}$ are collections of respective eigenvectors, $\mathbf{D}$ is a diagonal matrix with the singular values in its diagonal and n is the smallest dimension of $\mathbf{X}$ (or the number of nonzero singular values in the case where $\mathbf{X}$ is not full rank).

When EOFs are used to denoise the data, not all eigenvectors and singular values are used to reconstruct the data matrix. Instead, it is assumed that the eigenvectors corresponding to smaller singular values contain more noise with respect to the real data than the ones corresponding to larger values. Therefore, it is logical to select the q largest singular values and the corresponding vectors and reconstruct the denoised data matrix as:

$$\mathbf{X}_q = \sum_{k=1}^{q} \rho_k \mathbf{u}_k \mathbf{v}_k. \tag{1.9}$$

Brand [2002] proposed a recursive approach to estimate missing values. Since EOFs cannot be directly used with databases including missing values, those missing values must be replaced by some initial values in order to use the EOFs. A classical replacement is the mean value of the whole data matrix or the mean in one direction, row wise or column wise. The latter approach is more logical when the data matrix has some temporal or spatial structure in its columns or rows.

After the initial value replacement the SVD is done and the selected q singular values and eigenvectors are used to build the reconstruction. The initial missing values of $\mathbf{X}$ are replaced with the values from the reconstruction. After the replacement, a new SVD and reconstruction process are done again. The above procedure is repeated recursively until convergence criterion is fulfilled.

1.2.3 Robust Flexible Completion

The two methodologies previously presented are combined into the global methodology as presented hereafter. The RSOM algorithm for missing values is first ran. Then these first estimations serve as initialization for the EOFs method (the model calibration, number of code vector for SOM and singular values for EOFs, is done using validation

method).

As previously seen, SOM algorithm allows nonlinear projection. Code vectors will succeed to capture the nonlinear characteristics of presented inputs. However, SOM imputation is discrete: each time series belonging to the same cluster will have the same imputation candidate.

EOFs for missing values is very sensitive to initialization process. Because of the linearity of the EOFs approach, it will fail to reflect the nonlinear structures of the dataset. But contrary to SOM, EOFs remain continuous.

Due to these opposite characteristics, these two algorithms get along well together.

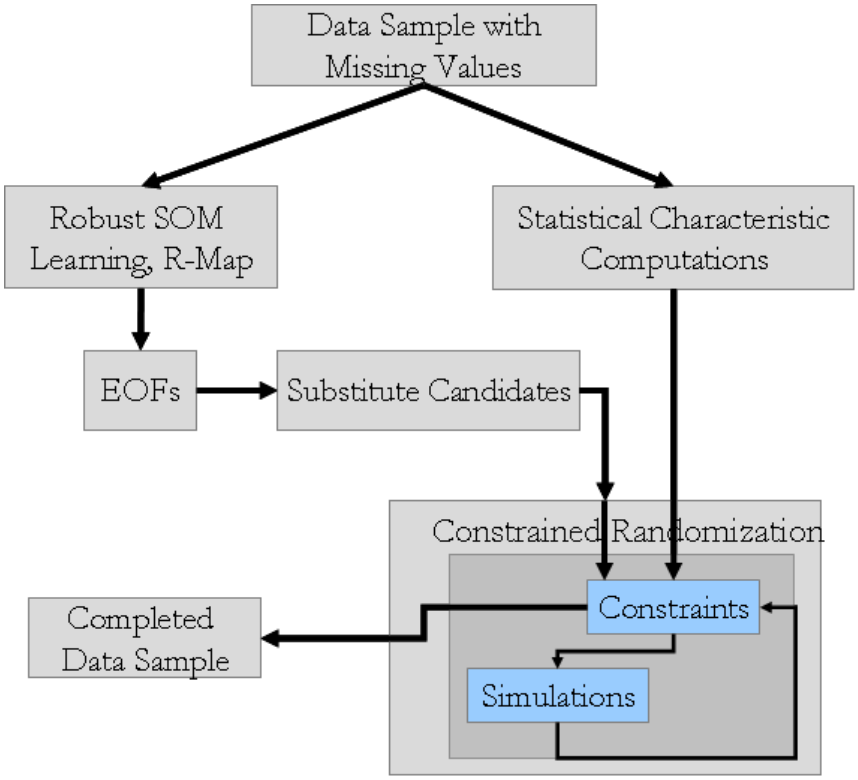
The above two-step procedure, when dealing with financial time series, will affect some important statistical properties of the over-all rebuilt dataset. In particular, higher moments (second, third and fourth centered moments), auto-correlations and the correlations with the other time series. We propose here to combine the SOM plus EOFs, adapted to the presence of missing values, and the Constrained Randomization algorithm introduced by Schreiber [1998]. This last computational method - initially presented as a specific reshuffling data sampling technique - allows to make a simulation of artificial time-series that fulfills given constraints, but are random in other aspects.

Figure 1.4 summarizes the proposed procedure for data completion, called Robust Flexible Completion (RFC). The first step starts with computing some empirical features of the data (moments and correlation of returns in our present case). Then, in parallel, a RSOM is run. Coordinates of Code Vectors in each of BMU are then considered as natural first candidates for missing value completion. This first estimation is used to initialize the EOFs procedure. Then, the constrained randomization, using as constraints some of the empirical features of the data determined at the first step, can start. If the candidate meets the constraints, then it takes the place of the missing value into the original data; if not, a residual noise[1] is drawn, then added to the previous candidates and the test of the constraints starts again. This process lasts until all constraints are fulfilled and all missing values replaced. We use the present values and their potential estimations done

[1]The noise is drawn from a central Skew Student's t-distribution introduced in Hansen [1994] with six degrees of freedom as mentioned in Patton [2004].

with RSOM plus EOFs to get the variance residual on the present values. According to these estimations, we fit the noise to draw for each time series.

Figure 1.4. Representation of the RFC



1.3 *Scenarii* Generation

When pure mathematical (closed form solution) analysis is not possible, simulations and *scenarii* provide a numerical alternative. From asset pricing to risk and performance measurement, simulations and *scenarii* generation are now widely spread in finance. We define, here, a *scenario* as a realization of a multivariate random variable (the return of each asset in our case). Number of methods exists to simulate return.

The simplest one is the historical approach. It is based on the assumption that past realisations are representative of future outcomes. The *scenarii* are usually considered as equally probable. The advantages of this method are that no assumption on the distribution of the asset returns is required. Indeed, the moment structures as well as the

dependancy of assets are inherent to the data. The main limits of such an approach are first, that the assumption regarding past realisations is somehow restrictive. It is reasonable to suppose that future returns differ significantly from the past ones. Secondly, the number of observations limits the number of *scenarii*.

To overcome to this last drawback, and produce a larger set of *scenarii*, it is common to use a bootstrap process (see Efron and Tibshirani [1993]). There are number of variant re-sampling techniques, but the general idea remains the same. From the original sample, a subsample is randomly drawn and compounded to get a new re-sample realization. Thus, bootstrap *scenarii* keep the advantages of the historical approach and allow to get the wanted level of outcomes. The only constraint is that the frequency of the original database has to be less than the horizon of *scenarii*. Since hedge fund return database frequencies are typically monthly, any bootstrap combination of return will lead several months horizon *scenarii*.

The remaining methods are parametric. The two widely spread approaches are the Monte Carlo simulation and the ARCH-based methods. The Monte Carlo simulations are based on specific distribution function assumptions. The choice of a Gaussian assumption has the kind advantage that the dependency structure is provided by the variance covariance matrix (see appendix C). In the case of rejection the Gaussian hypothesis, Monte Carlo simulation implementation is largely complexified (Embrechts, McNeil and Straumann [1999] and Embrechts [2008]). The dependency structure could be modelled with a copula (see Nelson [1997]). But this last approach suffers from the curse of dimensionality (see for instance, Mikosh [2006]) since the choice of the right copula becomes practically unrealizable.

We propose here to consider again the SOM algorithm. We have seen that SOM allows for classification of data samples with multiple variables; but when applied to longitudinal set, SOM can also classified multivariate realizations. Indeed, as pointed out by see Sarzeaud and Stéphan, [2000], non-linear interpolation can be realized with SOM. This last property is used to provide *scenarii* generation. The idea is to obtain a large map in which the subsets of returns at a particular date (the multivariate realizations) are

classified. Once the map has converged, each code vector represents a realistic realization of the multivariate return distribution. Considering large map, or multiple nested SOM classification allows us to generate the wanted number of *scenarii*. Such an approach is particularly useful when dealing with time series without enough historical data since historical simulation cannot be done.

1.4 Empirical Illustrations

The original data - provided by HFR^{TM} - consist in monthly Net Asset Values of Hedge Funds since January 85. From this sample, applying traditional filter rules, we keep 50 funds on a 10-year period. At the end, this one contains 50 funds and the number of observations is 120 (from December 1995 to December 2005). Note that, at purpose, no missing values are contained in this database.

Several funds included in the HFR^{TM} database use assets such as option and others exotic financial products inducing strong higher moment peculiarities. Furthermore, more than 10 different hedge funds strategies (see appendix 1) are present in this base leading to a very low global dependency (the mean correlation is .228, the mean Kendall's tau .160 and the mean Spearman's rho .226) and providing an interesting dataset to apply completion algorithm.

We artificially introduce various levels of missing data (from 5% to 25%) in our time series. We launch the imputation process hundred draws for each level of missing values. We implement the routines of the imputation methods in Matlab. The mean running time of excecution of the RFC is found to be above 5 minutes (45 seconds for SOM, 1 minute 15 seconds for EOF and 3 minutes and 15 seconds for the constrained randomization process) whereas ECM operates in less than 1 minute. But the RFC execution time could be reduced by limiting the constraints on the surrogate process depending of the wanted precision for the rebuilt process. Table 1.1 hereafter summarizes the mean properties of the errors in the first 4 moments and correlation when using the ECM algorithm and the two-step procedure presented in this article.

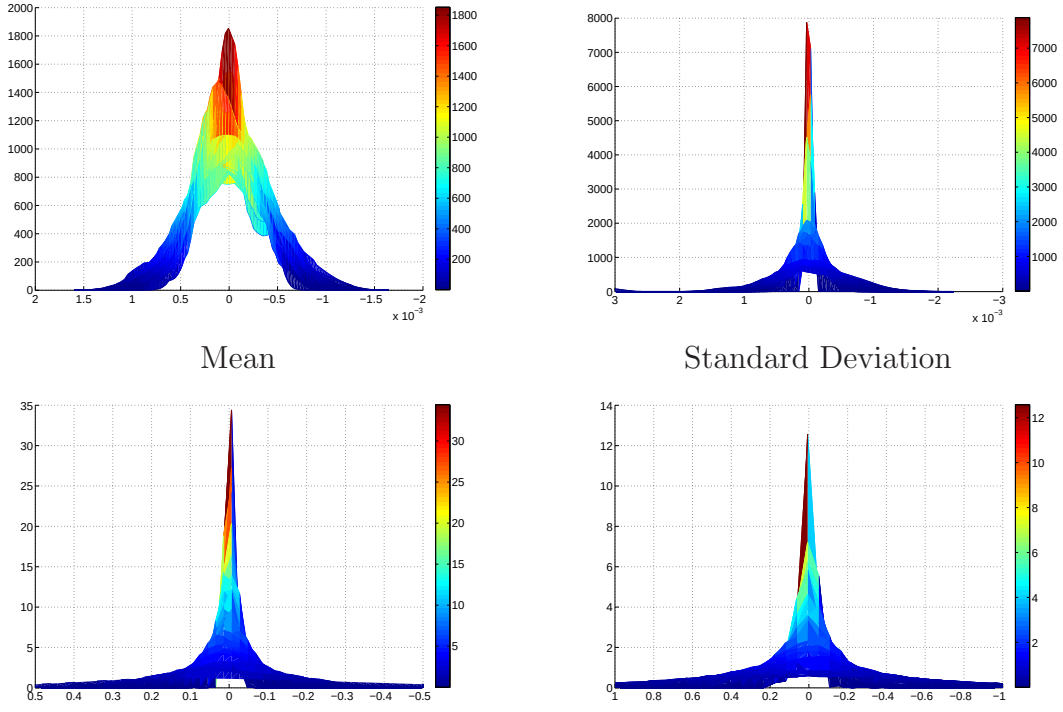
Table 1.1. Mean Errors on the Statistics of rebuilt Time Series.

Missing	**Squared Error after Completion**									
Values (in %)	**Mean**		**Std. Dev**		**Skewness**		*Kurtosis*		**Correlation**	
5	0.49	[5.07]	0.05	[0.12]	0.41	[4.49]	0.03	[0.87]	0.22	[0.24]
10	2.86	[14.03]	0.18	[0.28]	1.42	[18.29]	0.09	[3.99]	0.54	[0.57]
15	4.73	[23.17]	0.38	[0.40]	2.33	[20.19]	0.20	[4.82]	1.14	[0.81]
20	5.60	[44.47]	0.71	[0.41]	2.72	[22.29]	0.41	[3.87]	2.12	[1.17]
25	5.91	[33.46]	1.17	[0.43]	4.42	[35.70]	0.62	[8.67]	2.90	[1.05]

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Mean Errors on the statistics when using respectively the adapted RFC and the ECM (in brackets) algorithm for Missing Values (time $1\mathbf{E}^{-7}$,$1\mathbf{E}^{-5}$,$1\mathbf{E}^{-2}$,$1\mathbf{E}^{-7}$ respectively for mean, standard deviation, skewness and correlation) - for a hundred draws.

To illustrate the accuracy of the estimation procedure, the Figure 1.5 presents hereafter the non-parametric empirical densities of the centered errors of the first four moments for each fund obtained for hundred trials of the complete algorithm for a 25 % deletion level.

Figure 1.5. Error Densities of the First Four Moments.



Mean Standard Deviation

Skewness *Kurtosis*

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Representation of the densities of the first four moments of the 20 fund returns obtained for a 25% deletion level after hundred draws. Moments are on the x-axis, the different funds are on the y-axis, whilst the empirical estimations of the densities appear on the z-axis

1.4.1 Impact of Estimation Errors on Asset Allocation

For empirical illustrations, we choose practical examples. Since considering a fund of hedge fund, with 20 underlyings funds seems to be a reasonable assumption, we restrict our dataset to only 20 funds (following different hedge fund strategies) return time-series. We consider this subset of funds as our short list, and we perform asset allocations and efficient frontiers. On these 20 time series returns, we remove the first quarter of half of these time series (the 31 first returns). To allow an efficient imputation process of our algorithm, we add the 30 remaining complete returns times series. Thus the imputation process is done with 10% of missing values.

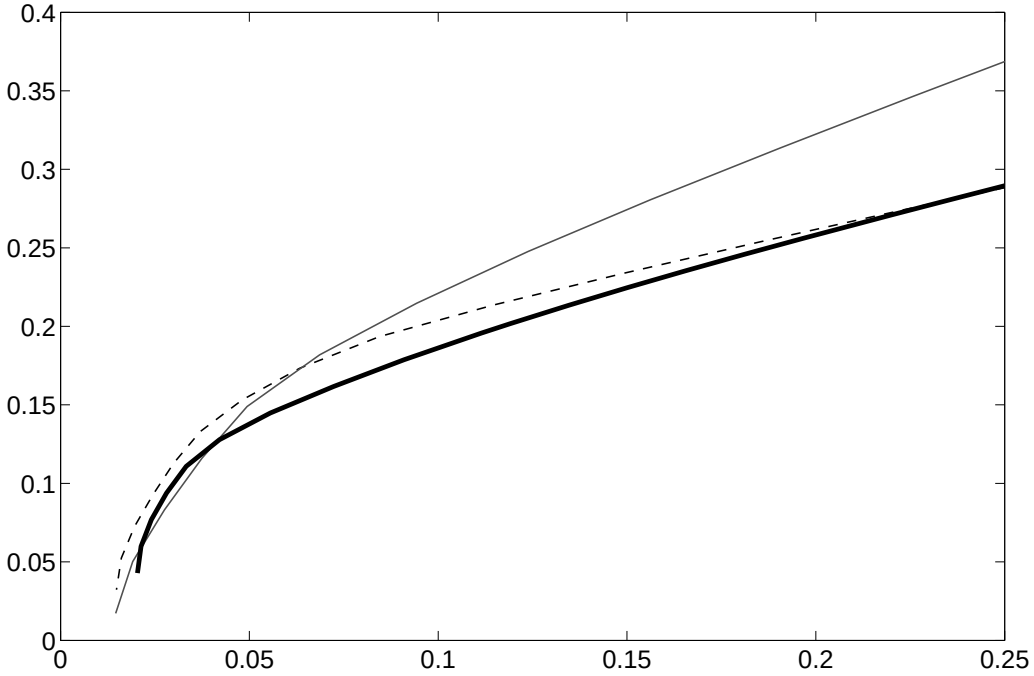
We consider the mean-variance framework proposed by Markowitz [1952]. We assume that the investor does not have access to a riskless asset, implying that the portfolio weights must sum to one. In addition we impose a no short-sale portfolio constraint: asset positions must be non-negative. Let $\mathbf{w}_p$ and $\mathbf{E}$ denote respectively the $(N \times 1)$ vector of weights and of expected returns for the N risky assets in the portfolio p; $\mathbf{\Omega}$ be the non-singular $(N \times N)$ variance-covariance matrix of the risky assets. Thus for a given level of expected return wanted $\mathbf{E}^*$, the optimal vector of weight $\mathbf{w}_{p^*}$ corresponding to the minimum variance portfolio selection problem is:

$$\begin{aligned} &\underset{\mathbf{w}_{p^*}}{Min}\, \mathbf{w}_p^{'} \mathbf{\Omega} \mathbf{w}_p \\ &s.t. \begin{cases} \mathbf{w}_p^{'} \mathbf{E} = \mathbf{E}^* \\ \mathbf{w}_{p^*}^{'} \mathbf{1} = 1 \\ \mathbf{w}_{p^*} \geq \mathbf{0} \end{cases} \end{aligned} \tag{1.10}$$

We perform this optimization problem for different levels of expected returns wanted, with the true complete 20 funds returns database, with RFC and the moments obtained with the ECM algorithm, see Figure 1.6.

Our estimation of the efficient frontier is clearly more in line with the true one. But since we focus on the asset allocation, we choose to highlight the composition of the maximum Sharpe portfolio. The weight associated to such a portfolio is as follow (with

Figure 1.6. Efficient Frontier after Completion.



Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Frontiers obtained after Missing Values Completions (for a 10% level of deletion). On the x-axis is the annualized volatility, on the y-axis the annualized expected returns. The dash grey line, continuous grey line and the bold continuous black line correspond respectively to efficient frontiers with true data, ECM completed data and RFC completed data.

previous notation):

$$\begin{aligned} &\underset{\mathbf{w}_{p^*}}{Max}\ \frac{\mathbf{w}_p^{'}\mathbf{E}-R_f}{\mathbf{w}_p^{'}\mathbf{\Omega}\mathbf{w}_p} \\ &s.t.\begin{cases} \mathbf{w}_{p^*}^{'}\mathbf{1}=1 \\ \mathbf{w}_{p^*}\geq\mathbf{0} \end{cases} \end{aligned} \tag{1.11}$$

where R_f is the risk free rate on the considered period.

1.4.2 Impact of Estimation Errors on Risks Measures

Different methods of Value-at-Risk have been developed. In the following section, we will briefly present the most used models (historical, parametric, semi-parametric, simulated).

Historical VaR corresponds to the quantile of return time series. On one hand, historical simulation is an attractive method because of the simplicity to implement and because it is a non parametric method (there is no need of assumptions about the underlying returns distribution). On the other hand, historical simulation is totally dependent of the selected period. A too short or unappropriated chosen period will lead to non accurate result. Therefore, historical simulation is a good proxy for more sophisticated VaR methods (subject to model risk).

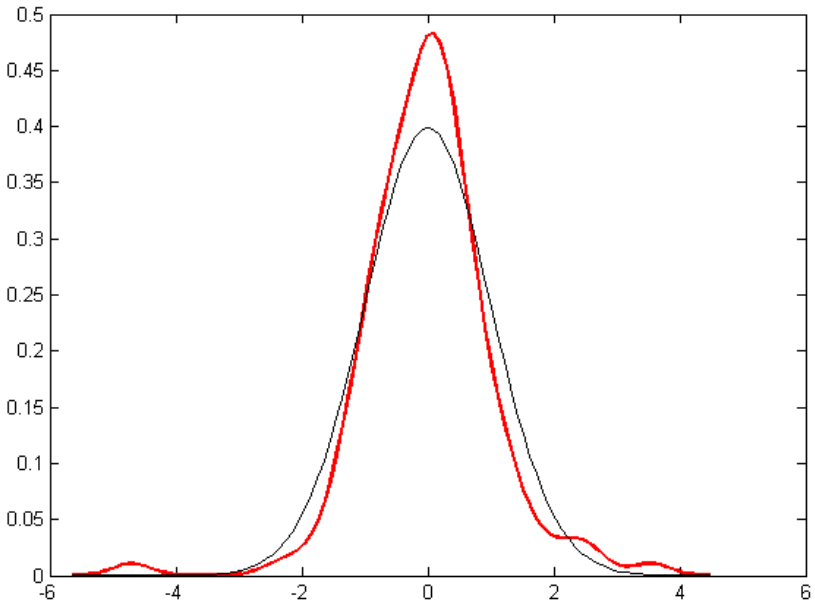
Parametric VaR focus on the cumulative density function (cdf) of returns. The two often used distributions are Gaussian and t-Student law. Gaussian assumption has the main advantage to be simple to estimate (only mean and standard deviation of returns need to be estimate) and remain easy to perform in a multidimensional framework (all the dependency structure is synthesized in the correlation matrix). Number of studies conclude that the normal assumption could be able to access VaR at a low confidence level (say 95%) but due to the fat tail or skew properties of many asset returns, the risk is underestimated when using a normal VaR. As an alternative the t-Student could be chosen to fit to empirical return. Thus, once the hypothesis of the distribution of returns have been set, the VaR at a wanted level α is simply:

$$VaR_{\alpha}(R) = \mathbf{E}(R) + F^{-1}(\alpha).\sigma(R) \tag{1.12}$$

where $\mathbf{E}(.)$ is the mean function and $\sigma(.)$ the standard deviation function and R the return time series.

Due to the stylized fact of hedge fund returns (non Gaussianity, positively skewed and leptokurtic return distribution), see Figure 1.7, we can hardly maintain Gaussian or t-Student return distribution hypothesis when dealing with hedge funds asset class. A

Figure 1.7. Density Estimation of HFR Global Index.



Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Comparison of empirical kernel density estimation of HFR Global Index returns with standard Gaussian distribution. The nonparametric density is estimated with a Gaussian kernel using the cross-validation criterion (*Cf.* Silverman, 1986

possible alternative consists of using semi-parametric VaR (see Favre and Galeano [2002]). Cornish-Fisher expansions (Cornish and Fisher [1937]), see Appendix B.1. It makes possible to adjust normal distribution to the true returns distribution according to skewness and excess *kurtosis*.

Table 1.2. Root Mean Square Error of Value-at-Risk obtained after Completion.

	Historical VaR		**Gaussien VaR**		**Cornish Fisher VaR**	
	ECM	**RFC**	**ECM**	**RFC**	**ECM**	**RFC**
VaR 95%	2.4%	2.2%	2.2%	1.5%	3.5%	2.1%
VaR 99%	7.3%	3.2%	3.1%	2.2%	12.4%	4.7%

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Root Mean Square Error of Value-at-Risk obtained after completion (for a 10% level of deletion) with ECM and RFC.

We perform the three VaR models at level 95% and 99% with the true return time series, and the two algorithms for missing values (see appendix C Table 1.5 and 1.6 for details). Table 2 shows the Root Mean Square Error (absolute) obtained with true data and those

obtained after recovery. Even for the Gaussian VaR which should be accurate with an ECM recovery, the SOM plus EOFs and constrained randomization perform better. The former algorithm permits to reduce estimation error of VaR measure from 20 to more than 700 basis points.

A last alternative to perform VaR when asset returns are non Gaussian, is the Monte Carlo simulation approach (MC VaR), see appendix B.2 for details. We choose to perform the MC VaR on two portfolios, the Maximum Sharpe exhibited previously (with weights obtained with the complete return time series) and the equaly-weighted portfolio. We also choose two distributions of risk factors, the Gaussian distribution, and a Student distribution with 6 degrees of freedom. The number of Monte Carlo iterations is set to 10 000 and we choose to measure VaR at a 99% and 95% thresholds.

Table 1.3. Absolute Error of Monte Carlo Value-at-Risk obtained for the Maximum Sharpe Portfolio.

	Gaussian Risk Factors		**Student Risk Factors**	
	ECM	**RFC**	**ECM**	**RFC**
VaR 95%	0.10%	0.03%	0.03%	0.10%
VaR 99%	0.14%	0.03%	0.07%	0.04%

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Absolute Error of Monte Carlo Value-at-Risk obtained for the maximum Sharpe portfolio with true returns and those obtained after completion (for a 10% level of deletion) with ECM and RFC.

Table 1.4. Absolute Error of Monte Carlo Value-at-Risk obtained for the equaly-weigthed portfolio.

	Gaussian Risk Factors		**Student Risk Factors**	
	ECM	**RFC**	**ECM**	**RFC**
VaR 95%	0.32%	0.10%	0.01%	0.01%
VaR 99%	0.37%	0.04%	0.06%	0.03%

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors. Absolute Error of Monte Carlo Value-at-Risk obtained for the equaly-weigthed portfolio with true returns and those obtained after completion (for a 10% level of deletion) with ECM and RFC.

Exhibit 10 shows the VaR measures obtained when applying MC VaR to the Sharpe portfolio and Exhibit 11 those obtained with the equaly weighted portfolio. Excepted for the case of MC VaR at 95% determination of Sharpe Portfolio with Student hypothesis

for the risk factor distribution, the risk measures obtained after RFC completion are more in line with the true one than those obtained after ECM reconstruction. More precisely, the mean absolute error on MC VaR is about 5 basis points when applying RFC and 14 basis points with ECM. Thus, RFC allows a reduction of absolute errors of 65%.

1.5 Conclusion

The presented method for data completion uses SOM description of the data as the starting point for EOFs to fill missing values. The first estimation is then slightly modified with a constrained randomization. The main interest of the technique can be found in the fact that some of the important empirical features of the input are respected during the rebuilding process of missing observations. Specifically higher moments and local correlations, whose accuracy of estimations are crucial in some financial applications, are taken into account when substitutions are performed. Moreover, one can easily think about some generalizations of the submitted algorithm, adding for instance some features under studies into the constraints of the so-called Constrained Randomization procedure, such as auto-correlation structure, depending on what is the final aim of the financial applications (asset allocation or risk management, see Jondeau and Rockinger, 2003). One also can think to modify RFC to perform *scenario* generation and prediction.

1.6 Appendix

1.6.1 HFR^{TM} Main Classification - Strategies Definitions

- Global Macro funds aim to profit from changes in global economies as influenced by major economic trends and/or events. They use leverage and derivatives to accentuate the impact of market moves. In consequence their expected volatilities are very high.

- Emerging market hedge funds invest in equity or debt of emerging markets. There are no viable futures or other derivative products for hedging. Emerging market

hedge funds can be partially hedge via U.S. treasury futures and currency markets, but their expected volatility is very high.

- Fund of Funds mixes and matches hedge funds. This blending of different strategies and asset classes aims to provide a high level of diversification and a more stable long-term investment return than any of the individual funds. Returns, risk, and volatility can be controlled by the mix of underlying strategies and funds.

- Market neutral hedge funds tend to negate the impact and risk of general market movement. There are two principal sub-categories: Market neutral Arbitrage (attempts to hedge out most market risk by taking offsetting positions) and Market Neutral Securities Hedging (invests equally in long and short equity portfolios generally in the same sectors of the market). Due to the deep level of exposure to the stock market, the expected volatility is generally low.

- Convertible arbitrage hedge funds entail buying a corporate convertible bond, while simultaneously selling short the common stock of the same company that issued the bond. The idea is to make money from the bond's yield if the bond goes up but also make money from the short sale if the stock goes down. As the convertible bond and the stock can move independently, this investment is very risky. The expected volatility is then high.

- Fixed income arbitrage hedge funds tend to exploit pricing inefficiencies between fixed income securities and neutralize exposure to interest rate risk. This type of fund often exploits arbitrage opportunities in interest rate swaps, government bonds, nongovernmental bonds, forward agreements.

- Event-driven hedge funds focus on price movements observed during anticipation of corporate event such as leveraged buy-outs, mergers and hostile takeovers. The most common event driven strategies are distressed securities and merger arbitrage.

- Long/short equity hedge funds combine long positions with short sales. For example, a long/short manager might purchase a portfolio of core stocks that occupy the S&P

500 and hedge by selling (shorting) S&P 500 Index futures. If the S&P 500 goes down, the short position will offset the losses in the core portfolio, limiting overall losses.

- Dedicated short bias hedge funds are specialized in the short sale of over-valued securities. Because losses on short-only positions are theoretically unlimited (because the stock can rise indefinitely), these strategies are particularly risky.

- Managed Futures funds invest in the global currency, interest rate, equity, metal, energy and agricultural markets. They do this through the use of futures, forwards and options.

1.6.2 Expected Maximization Methods

As benchmark to estimate how well perform our two steps procedure, we choose to compare our results with those obtained by the EM algorithm.

The Expectation Maximization (EM) algorithm presented in Dempster, Laird and Rubin [1977] is a technic to find the maximum likelihood estimates in missing data situation. Since estimates of the mean and of the covariance matrix of an incomplete dataset depend on the unknown missing values, and since, conversely, estimates of the missing values depend on the unknown statistics of the data, this estimation problem is a non linear one and has to be done iteratively.

The EM algorithm consists of two steps:

1. E-step calculates the expectation of the complete data sufficient statistics given the observed data and current parameter estimates.

2. M-step updates the parameter estimates through the maximum likelihood approach based on the current values of the complete sufficient statistics.

The algorithm then proceeds in an iterative manner until the difference between the last two consecutive parameter estimates converges to a specified criterion. The final E-

step calculates the expectation of each missing value given the final parameter estimates and the observed data; this will be used as the imputation value.

For each iteration (t), the E-step consist to calculate:

$$Q\left(\theta \left| \theta^{(t)} \right.\right) = E\left[L\left(\theta \left| Y \right.\right) \left| Y_{obs}, \theta^{(t)} \right.\right], \tag{1.13}$$

where

$$\begin{cases} L\left(. \left| Y \right.\right) \text{ denotes the likelihood function conditionally to the sample,} \\ \theta \text{ the vector of parameter to be estimated,} \\ Y_{obs} \text{ the non missing values,} \\ Y \text{ the sample,} \\ \theta^{(t)} \text{ the last vector of parameter estimated.} \end{cases}$$

Then the $(t+1)^{th}$ M-step finds $\theta^{(t+1)}$ to maximize $Q\left(\theta \left| \theta^{(t)} \right.\right)$ such that:

$$Q\left(\theta^{(t+1)} \left| \theta^{(t)} \right.\right) = \max_{\theta} Q\left(\theta \left| \theta^{(t)} \right.\right). \tag{1.14}$$

The main drawback of the EM algorithm is when the M-step is not in close form. In this case, the M-step could be difficult to perform. Meng and Rubin [1993] proposed an alternative algorithm called the Expectation Conditional Maximization (ECM) to solve this problem. The M-step is decomposed in multiple conditional maximization. Let consider $\theta = [\theta_1, \theta_2, ..., \theta_k]$ a k-dimensional vector of parameters. Then the CM-step consists in k successive maximizations, with previous notation:

$$Q\left(\theta^{(t+1)} \left| \theta^{(t)} \right.\right) = \max_{\theta_i} Q\left(\theta \left| \theta^{(t)} \right.\right), \text{ for } i = 1, ..., k. \tag{1.15}$$

1.6.3 VaR Methodologies

Brut Simulation Methods

Simulation of multivariate variable is a classifical problem. Under the Gausian assumption, we only have to focus on the two first empirical moments as well as the correlation stucture. Cholesky decomposition allows to easily perform such a simulation. Let's consider a multivariate Gaussian variable X, with mean vector μ, and a covariance matrix $\mathbf{\Sigma}$. Compute the Cholesky decomposition (matrix square root) of $\mathbf{\Sigma}$ is to find the unique lower triangular matrix $\mathbf{A}$ such that:

$$\mathbf{A}\mathbf{A}'=\mathbf{\Sigma}. \tag{1.16}$$

To generate a realization $\mathbf{X}$ of X, we just need to draw $\mathbf{Z}$ a vector whose component are independant standard variates

$$\mathbf{X}=\mu+\mathbf{A}\mathbf{Z}. \tag{1.17}$$

Once we have generated futur code vector values, we are now able, applying the surrogate data procedure describe below to rebuild the futur values of funds returns.

Cornish-Fisher VaR

Cornish-Fisher expansions provide an elegant way to integrate the higher moments of distribution. Cornish-Fisher expansions (Cornish and Fisher [1937]) permit to adjust normal distribution to the true returns distribution according to skewness and excess *kurtosis*:

$$f(r)=\left\{1+\frac{\gamma_1[f(r)]}{3!}H_2(r)+\frac{\gamma_2[f(r)]}{4!}H_3(r)+10\frac{\gamma_1[f(r)]^2}{36}H_6(r)\right\}\phi(r) \tag{1.18}$$

where $\phi(r)$ is the standard normal probability density function, γ_1 is the skewness and γ_2 is the excess *kurtosis* and $H_i(r) = (-1)^i \phi(r)^{-1} \left[\frac{\partial^i \phi(r)}{\partial r^i}\right]$ the Hermite polynomial at order i.

Thus the quantile at level α (denotes Q_α) is then:

$$\begin{array}{c} Q_\alpha[f(r)] = Q_\alpha[\Phi(r)] + \frac{\gamma_1[f(r)]}{3!} H_2\{Q_\alpha[\Phi(r)]\} + \frac{\gamma_2[f(r)]}{4!} H_3\{Q_\alpha[\Phi(r)]\} \\ -\frac{\gamma_1[f(r)]^2}{36}\left\{2Q_\alpha[\Phi(r)]^3 - 5Q_\alpha[\Phi(r)]\right\} \end{array} \tag{1.19}$$

where $\Phi(r)$ is the standard normal probability density function

Thus the Cornish-Fisher VaR (see Zangari, [1996]) at level α is then (with previous notation):

$$VaR_\alpha(R) = \mathbf{E}(R) + Q_\alpha[f(r)].\sigma(R) \tag{1.20}$$

PCA VaR

The PCA allows to find an equivalent input space where the risk factors are non correlated. This is done through linear transformation. PCA can be calculated by solving the following eigenvalue problem:

$$\mathbf{P}^{-1}\mathbf{\Omega}\mathbf{P} = \mathbf{D} \tag{1.21}$$

where $\mathbf{\Omega}$ is the variance covariance matrix associated to our data $\mathbf{X}$, $\mathbf{D}$ is a diagonal matrix containing the singular values and $\mathbf{P}$ is transformation matrix performing the linear transformation.

This $\mathbf{P}$ matrix is then used to obtain the time series of the non correlated risk factor returns.

$$\mathbf{R} = \mathbf{X}\mathbf{P} \tag{1.22}$$

After having chosen the distribution of risk factors, we apply the Monte Carlo simulation to get $\mathbf{R}_{simu}$ the matrix of realistic risk factor realizations. Then we apply the reverse linear transformation by multiplying $\mathbf{P}^{-1}$ to $\mathbf{R}_{simu}$, we come back to the real input space

and get $\mathbf{X}_{simu}$ the matrix of realistic return realizations.

$$\mathbf{X}_{simu} = \mathbf{R}_{simu}\mathbf{P}^{-1} \tag{1.23}$$

Once we have $\mathbf{X}_{simu}$, we only need to apply the weight of the considered portfolio to the simulated returns and using the appropriated quantile at the wanted threshold, to the obtained series, to get the MC VaR.

1.6.4 Value-at-Risk Estimation

Table 1.5. Value-at-Risk 99%.

	Historical VaR			Gaussian VaR			Cornish Fisher VaR		
	True	ECM	RFC	True	ECM	RFC	True	ECM	RFC
Funds 1	-7.0%	-4.6%	-4.6.%	-3.2%	-4.7%	-3.0%	-8.5%	-3.8%	-7.0%
Funds 2	-3.3%	-3.9%	-2.7%	-2.4%	-3.3%	-2.4%	-2.4%	-2.7%	-2.2%
Funds 3	-6.4%	-6.1%	-5.0%	-6.0%	-6.8%	-5.3%	-6.3%	-6.1%	-5.2%
Funds 4	-6.1%	-7.0%	-6.1%	-2.1%	-3.0%	-2.2%	-5.2%	-5.6%	-5.2%
Funds 5	-8.5%	-8.0%	-8.0%	-9.7%	-13.1%	-9.3%	-8.3%	-6.6%	-7.3%
Funds 6	-3.1%	-5.9%	-2.6%	-2.6%	-3.7%	-2.0%	-3.4%	-6.2%	-2.3%
Funds 7	-39.6%	-30.4%	-29.0%	-32.6%	-35.4%	-24.0%	-43.3%	-23.5%	-25.4%
Funds 8	-17.5%	-18.4%	-17.5%	-13.0%	-15.0%	-12.7%	-18.1%	-19.6%	-20.5%
Funds 9	-4.8%	-6.9%	-4.4%	-3.1%	-3.5%	-2.9%	-4.6%	-6.3%	-4.5%
Funds 10	-5.7%	-9.7%	-5.1%	-4.9%	-6.1%	-4.5%	-4.6.%	-10.5%	-4.9%
Funds 11	-6.9%	-10.6%	-7.6%	-6.6%	-8.3%	-5.8%	-6.5%	-9.8%	-7.3%
Funds 12	-14.8%	-43.8%	-14.8%	-14.1%	-25.1%	-13.6%	-11.9%	-58.3%	-13.5%
Funds 13	-2.3%	-7.7%	-3.6%	-3.5%	-5.4%	-3.9%	-2.8%	-8.3%	-3.2%
Funds 14	-4.2%	-4.2%	-4.2%	-3.9%	-4.4%	-3.9%	-3.3%	-3.1%	-3.6%
Funds 15	-10.7%	-15.8%	-10.7%	-11.5%	-15.6%	-10.9%	-9.9%	-13.2%	-10.3%
Funds 16	-.8%	-1.5%	-1.2%	-1.0%	-1.3%	-1.4%	.2%	-1.4%	.5%
Funds 17	-13.9%	-13.9%	-13.9%	-11.0%	-10.1%	-8.9%	-12.9%	-12.4%	-12.8%
Funds 18	-13.0%	-15.3%	-12.6%	-10.4%	-11.3%	-10.0%	-12.2%	-16.1%	-11.8%
Funds 19	-5.6%	-5.6%	-5.6%	-1.9%	-3.1%	-2.1%	-5.7%	-5.5%	-5.7%
Funds 20	-21.7%	-26.5%	-12.4%	-15.2%	-18.6%	-11.8%	-21.3%	-40.7%	-11.3%

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors.

Table 1.6. Value-at-Risk 95%.

	Historical VaR			Gaussian VaR			Cornish Fisher VaR		
	True	ECM	RFC	True	ECM	RFC	True	ECM	RFC
Funds 1	-1.1%	-2.7%	-1.1.%	-2.1%	-3.0%	-1.9%	-1.9%	-2.0%	-1.5%
Funds 2	-1.1%	-1.6%	-1.2%	-1.5%	-2.1%	-1.5%	-1.1%	-1.5%	-1.0%
Funds 3	-4.2%	-4.2%	-3.9%	-3.9%	-4.5%	-3.4%	-4.1%	-4.1%	-3.5%
Funds 4	-1.0%	-1.7%	-1.2%	-1.4%	-2.1%	-1.4%	-1.8%	-2.6%	-1.8%
Funds 5	-6.5%	-6.5%	-5.0%	-6.7%	-8.8%	-6.4%	-6.1%	-2.3%	-5.6%
Funds 6	-1.7%	-3.2%	-1.3%	-1.6%	-2.5%	-1.2%	-1.1%	-3.1%	-1.2%
Funds 7	-24.9%	-21.8%	-15.7%	-22.6%	-23.4%	-16.0%	-21.2%	-17.0%	-12.8%
Funds 8	-8.3%	-12.2%	-8.5%	-8.9%	-10.2%	-8.6%	-5.0%	-8.1%	-5.3%
Funds 9	-1.7%	-2.0%	-1.5%	-2.0%	-2.4%	-1.9%	-2.2%	-2.8%	-2.1%
Funds 10	-2.6%	-2.1%	-1.9%	-3.2%	-4.0%	-2.9%	-2.1.%	-3.6%	-1.6%
Funds 11	-4.2%	-6.8%	-3.5%	-4.3%	-5.8%	-3.8%	-4.3%	-6.2%	-4.0%
Funds 12	-7.9%	-14.9%	-8.7%	-9.5%	-17.6%	-9.2%	-7.7%	-20.1%	-8.0%
Funds 13	-2.0%	-3.3%	-2.0%	-2.0%	-3.6%	-2.4%	-1.4%	-3.6%	-1.7%
Funds 14	-2.2%	-2.3%	-2.2%	-2.5%	-2.7%	-2.5%	-1.2%	-1.8%	-1.3%
Funds 15	-6.1%	-10.3%	-6.1%	-8.0%	-10.8%	-7.5%	-7.0%	-9.7%	-6.9%
Funds 16	-.1%	-1.0%	-.4%	-.7%	-.9%	-.9%	-.2%	-.6%	.2%
Funds 17	-9.4%	-7.3%	-6.1%	-7.6%	-6.9%	-6.0%	-7.7%	-7.0%	-6.1%
Funds 18	-7.3%	-7.3%	-6.5%	-7.1%	-7.6%	-6.7%	-7.7%	-8.6%	-7.3%
Funds 19	-.2%	-1.2%	-.5%	-1.0%	-1.9%	-1.2%	-1.4%	-2.0%	-1.6%
Funds 20	-8.0%	-10.6%	-7.0%	-10.4%	-13.2%	-7.9%	-9.7%	-14.5%	-6.4%

Source: HFR^{TM}; Monthly Net Asset Values (12/1995-12/2005). Computations from the authors.

Chapter 2

Efficient Frontier for Robust Higher-order Moment Portfolio Selection

2.1 Introduction

Since the first mention of higher-order moments than the variance of returns by Marschak (1938) and Hicks (1939), it is now generally accepted by the financial community that investors generally exhibit preferences for positively skewed and light-tailed asset return distributions (see, for instance, Beedles and Simkowitz (1978), Dittmar (2002), Jurczenko and Maillet (2006a), and Mitton and Vorkink (2007)). Developments with higher moments followed since the origin three main (complementary) directions in finance: a tentative integration in the von Neumann-Morgenstern utility function of a rational investor (*Cf.* Arrow (1964) and Pratt (1964)) of higher-order moments of returns by Arditti (1967), Samuelson (1970) and Tsiang (1972); an attempt to generalize the Markowitz (1952) efficient frontier to incorporate the effect of higher moments on optimal asset allocations by Jean (1971 and 1973), Arditti and Levy (1972), Ingersoll (1975) and Schweser (1978); and a first partial explicit modelling of returns by Rubinstein (1973), and Kraus and Litzenberger (1976), through extensions of the CAPM by Sharpe (1964). True depar-

tures from Gaussianity may indeed affect the optimal allocation of assets (see Jondeau and Rockinger (2003b and 2006)) and the mean-variance portfolio selection *criterion* proposed by Markowitz (1952) is *a priori* somehow inadequate for some risky assets whose characteristics are very special (see Jondeau and Rockinger (2006)). In such a context, different multi-moment approaches have been proposed in the financial literature to incorporate higher-order moment preferences into asset allocation problems[1], in order to characterize generalized geometric efficient frontiers (see Athayde and Flôres (2002 and 2003)); but all suffer more or less from the traditional drawbacks of algebraic moments. First, higher-order moments do not always exist (see Embrechts, Kluepelberg and Mikosh (1997), and Jondeau and Rockinger (2003a and 2003b)), and even when they do, such moments do not always uniquely define a probability distribution, so that two distinct distributions can have the same sequence of moments (see Heyde (1963) for the example of the Log-normal density). Secondly, conventional moments tend to be very sensitive to a few extreme observations (see Hampel, Ronchetti, Rousseeuw and Stahel (2005)). The asymptotic efficiency of the empirical moments is also rather poor, especially for distributions with fat tails. This last property is an immediate consequence of the fact that the asymptotic variances of these estimators are mainly determined by higher-order moments, which tend to be rather large, or even unbounded, for heavy-tailed distributions.

The objective of this paper is to overcome the limits of traditional multi-moment asset allocation models by using an alternative set of statistics in traditional optimization programs, namely L-moments. Recent attempts for modelling distributions in a multivariate framework are indeed built on the concept of order-statistics, for calibrating a Bernstein *Copula* in Baker (2008) or for defining extreme co-movements using L-moments in Serfling and Xiao (2007). The latter, which are linear functions of the expectations of order statistics, were introduced under this name by Sillitto (1951) and comprehensively

[1]See Athayde and Flôres (2002 and 2006), Briec, Kerstens and Jokung (2007), Jurczenko and Maillet (2006b), and Jurczenko, Maillet and Merlin (2006) for the primal approaches of the mean-variance-skewness-*kurtosis* portfolio decision problems; and Simaan (1993), Gamba and Rossi (1998a and 1998b), Jurczenko and Maillet (2001 and 2006a), and Jondeau and Rockinger (2006) for the dual approaches of the higher-order moment asset allocation problems.

reviewed by Hosking (1989). As so-called U-statistics (see Hoeffding (1948)), L-moments offer one main advantage over Conventional moments (denoted herein C-moments as in Ulrych, Velis, Woodbury and Sacchi (2000), and Chu and Salmon (2008)). Their empirical counterparts are less sensitive to the effects of sampling variability, since they are linear functions of the ordered data, and are therefore shown to provide more robust estimators of higher moments than the corresponding sample C-moments (Sankarasubramanian and Srinivasan (1999)). More precisely, L-moments are defined as certain linear functions of the Probability Weighted Moments (Greenwood, Landwehr, Matalas and Wallis (1979)) and can characterize a wider range of distributions compared to C-moments. Indeed, they exist whenever the mean of the distribution does, even though some C-moments do not (which is very likely to be the case in finance). As we will see later on, they are also particularly well adapted for addressing some specific concerns in the field of finance. And the beauty is that they are easy and fast(er) to compute, besides being reliable estimators of characteristic shape parameters of general distributions.

Because of their proven advantages, L-moments have already found wide applications in various fields such as meteorology, hydrology, geophysics and regional analysis (see Hosking and Wallis (1997)) - where large deviations really matter, namely for instance when studying extreme floods or low flows (see among the main references: Hosking and Wallis (1987), Ben-Zvi and Azmon (1997), Wang (1997), Bayazit and Önöz (2002), Moisello (2007), and Shao, Chen and Zhang (2008)), rainfall extremes (see Guttman, Hosking and Wallis (1993), Lee and Maeng (2003), and Parida and Moalafhi (2008)), raindrop sizes (see Kliche, Smith and Johnson (2008)), velocity of gale force winds (see Pandey, Gelder and Vrijling (2001), Whalen, Savage and Jeong (2004), and Modarres (2008)) or the measurement of earthquake intensities (Thompson, Baise and Vogel (2007)). Lately, they have also found an interest in finance, first, as *Monsieur* Jourdain by Molière without noticing it, when using a special case of an L-moment which is the Gini coefficient (see Gini (1912) and below) as a substitute to the volatility in asset pricing models (see Shalit and Yitzhaki (1989), Okunev (1990), and Benson, Faff and Pope (2003)); then,

secondly, for more general purposes: for fitting return distributions (Hosking, Bonti and Siegel (2000), Carrillo, Hernández and Seco (2006a), and Karvanen (2006)) and the rate of profit densities (Wells (2007)), the design of a GMM-type Goodness-of-Fit test (see Chu and Salmon (2008)), risk modelling purposes (see Martins-Filho and Yao (2006), Tolikas and Brown (2006), Tolikas, Koulakiotis and Brown (2007), Gouriéroux and Jasiak (2008), and Tolikas (2008)), calibrating extreme return distributions (Gettinby, Sinclair, Power and Brown (2006), French (2008), and Tolikas and Gettinby (2009)) or rogue-volatility densities (Maillet and Médecin (2008), and Maillet Médecin and Michel (2008)), and very recently for defining a new set of measures of performance for hedge funds (see Darolles, Gouriéroux and Jasiak (2008)).

Thanks to the so-called shortage function technique and relying on robust L-statistics, we thus generalize in this article the traditional mean-variance-skewness-*kurtosis* efficient frontier in the four L-moment space, proposing a new and fast formulation of higher-order (L-)comoments of efficient portfolios. The shortage function of Luenberger (1995) was indeed first applied to the portfolio performance evaluation in the traditional mean-variance framework by Morey and Morey (1999), then developed by Briec, Kerstens and Lesourd (2004), and recently extended to multi-horizon performance appraisals (Briec and Kerstens (2009)). In brief, the shortage function rates the performance of any portfolio by measuring a distance between the coordinates of this specific portfolio and those of its radial projection onto the multi-moment efficient frontier. Based on this distance definition, the so-called Goal Attainment Method enables us to solve the multiple conflicting and competing allocation objectives, without assuming a detailed knowledge of the preference parameters of the indirect investor's utility function.

After a theoretical presentation of L-moments and of the shortage function approach in a portfolio selection context, we propose to rewrite the multi-moment optimization program of the investor within a new compact notation, using both C-moments and L-moments, then derive the four L-moment efficient set and provide various illustrations

using a universe of 162 European stocks. Our empirical results regarding links between moments of efficient portfolios and the various shapes of the higher-order moment efficient frontier - from (*pseudo-*)*parabolae* to (deformed) cones, confirm the earlier findings by Jean (1973) and Ingersoll (1975) presenting the first three-dimensional representation of the efficient set. They are also consistent with more recent evidences provided in other frameworks (see for instance, Athayde and Flôres (2004), Jurczenko, Maillet and Merlin (2006), Maringer and Parpas (2008)). However, it is worth noting that our attempt to evaluate the cost of not using higher-order moments is still not conclusive in a traditional "mixed" utility setting, since differences between optimal asset allocation implied utilities are found to be marginal. In other words, the mean-variance *criterion* - corner stone of the Modern Portfolio Theory of de Finetti[2] (1940) and Markowitz (1952) - is shown to be rather accurate as predicted, among others, by Levy and Markowitz (1979) and Kroll, Levy and Markowitz (1984). As mentioned by Jondeau and Rockinger (2006), the use of higher-order moments, in a the traditional expected utility framework and in a restrictive "mixed" utility setting (in which sensitivities only depends upon the first moment), may thus only prove their efficiency either if the underlying assets are largely non-Gaussian (in some specific sense) or if the representative investor exhibits some very peculiar features regarding her prudence and temperance characteristics. Nevertheless, we also suggest that complementary controls of higher-order moments of some traditional low-dispersion return efficient portfolios (such as the Global Minimum-Variance Portfolio) may be of interest for the investors.

The remainder of the article is organized as follows. In section 2, we formally present the L-moments, briefly recall some of their main properties, and illustrate their computations on a long record of stock index quotes. In section 3, we precisely define in a new notation higher-order moments of returns on portfolios and describe how the optimal portfolio selection L-moment program can be solved in a shortage function framework. In section 4, we present the data and discuss the results of the various optimal asset al-

[2]See Markowitz (2006), Pressacco and Serafini (2007) and Barone (2008).

locations. Section 5 concludes. The Appendices are dedicated to proofs, some technical details, Tables and Figures.

2.2 Robust Higher-order Moments for Portfolio Selection

Introduced by Sillitto (1951) and popularized by Hosking (1989), L-moments can be interpreted, like C-moments, as simple descriptors of the shape of a general distribution, albeit offering a number of advantages.

First, all population (higher) L-moments exist and uniquely determine a probability distribution, provided that the mean exists (see Chan (1967), and Arnold and Meelden (1975)). In this case, a distribution can always be specified by its L-moments, even if some of its higher-order C-moments do not exist. Furthermore, this specification is always unique. For the standard errors of L-moments to be finite, it is also only required that the distribution has a finite variance; no condition on higher-order moments is necessary (Hosking (1990)). Moreover, although moment ratios can be arbitrarily large, sample moment ratios have algebraic bounds (see Dalén (1987)) and sample L-moment ratios can take any values that the corresponding population quantities can (Hosking (1990)). Motivated by the sampling properties of L-statistics, Hosking and Wallis (1987) thus advocate that L-moments provide a better approximation of the unknown parent distribution than C-moments. They also provide reliable estimators for Extreme Value densities and have been widely used in fields where exceptions are at stake (see Hosking (1990)).

Secondly, L-moments exhibit some specifically interesting features for financial applications. Since a (complete) set of L-moments determine a unique density, the so-called Hamburger problem (see Jondeau and Rockinger (2003a) and Jurczenko and Maillet (2006a)) - when C-moments lead to several laws - is limited. Since they always exist, the problem of working with non-defined quantities such as higher C-moments is avoided. L-moments are also coherent shape measures of risk (see Artzner, Delbaen, Eber and Heath (1999)), since they are translation and scale invariants (Serfling and Xiao (2007), and Gouriéroux

and Jasiak (2008)). They, furthermore, allow us a clearer focus on a specific part of the distribution, thus avoiding confusion between the center and the extreme parts of the distribution as in the traditional case (see Haas (2007)). Finally, the sample estimates of L-moments are more robust to data outliers (Vogel and Fennessey (1993)) - since they are only linearly influenced by large deviations (see Hosking (1990)) - and more efficient than C-moments (see Sankarasubramanian and Srinivasan (1999), and Carrillo, Hernández and Seco (2006b)), especially within a Generalized Method of (L-)Moment context (see Gettinby, Sinclair, Power and Brown (2006), Chu and Salmon (2008), and also Gouriéroux and Jasiak (2008)).

After having briefly recalled the different analytical representations of the univariate population L-moments[3] - we mainly refer here to the work by Hosking and Wallis (1997), we will present in the following sub-section their sample unbiased estimator counterparts and shall illustrate the four first L-moment estimates on a long sample of one century of daily quotes of the Dow Jones Index.

2.2.1 Population L-comoments

Population L-moments are defined as certain linear functions of the expectations of the order statistics from the population distribution of the underlying random variable. Let us start with some basic notations and definitions.

Let $\{X_t\}$, with $t = [1, ..., T]$, be a conceptual random sample of size T drawn from a continuous probability distribution $F(.)$ of a real-valued random variable X, with $T \in IN^*$; $Q(u) = F^{-1}(u)$, for $u =]0, 1[$, a quantile function, and $X_{[1:T]} \leq X_{[2:T]} \leq ... \leq X_{[T:T]}$ denoting the corresponding order statistics. Then the k-th population univariate L-moment (using the L-functional representation) is defined, $\forall k \in IN^*$ and $k < T$, as

[3]Other variants of L-moments, called TL-moments (encompassing the PL-moments, the LH-moments and the LL-moments), LQ-moments (see Maillet and Médecin (2008), for a comprehensive review and an application to extreme volatilities), as well as the LSD-moments (SD-PWM - see Haktanir (1997), and Whalen, Savage and Jeong (2004)), have already been used in extreme studies, since they are proven to be even less sensitive to outliers. However, to be as close as possible to the portfolio value as perceived by the investors, we will stick, in this article, with the traditional simple L-moments in our portfolio context.

(see, for instance, Hosking (1990)):

$$\begin{array}{rcl} \lambda_k(X) & = & \sum_{j=0}^{k-1}(-1)^j\left\{[k\ (k-j-1)!j!]^{-1}[(k-1)!]\right\}\times E\left(X_{[k-j:k]}\right) \\ & = & \int_0^1 Q(u)\ P_{k-1}^*(u)\ du, \end{array} \tag{2.1}$$

with:

$$\left\{\begin{array}{rcl} E\left(X_{[r:k]}\right) & = & \left\{[(r-1)!(k-r)!]^{-1}(k!)\right\}\times\int_0^1 Q(u)\,u^{r-1}(1-u)^{k-r}\,du \\ P_k^*(u) & = & \sum_{r=0}^{k} p_{k,r}^*\,u^r \\ p_{k,r}^* & = & (-1)^{k-r}\left\{\left[(r!)^2(k-r)!\right]^{-1}(k+r)!\right\}, \end{array}\right.$$

where $\lambda_k(.)$ is the L-moment of order k, $r=[1,...,k]$, $0<u<1$, $E(.)$ is the expectation operator and $p_{k,r}^*$ corresponds to the r-th coefficient of the shifted orthogonal Legendre polynomial of degree k denoted $P_k^*(.)$ and defined as $P_k^*(u)=P_k(2u-1)$ where $P_k(.)$ is the traditional Legendre polynomial of degree k.

Thus, the shifted orthogonal Legendre polynomial satisfies, $\forall(k,s)\in IN^{*2}$:

$$\left\{\begin{array}{ll} \int_0^1 P_r^*(u)\ P_s^*(u)\ du=(2r-1)^{-1} & \text{if } r=s \\ =0 & \text{if } r\neq s, \end{array}\right. \tag{2.2}$$

with $P_0^*(u)=1$.

In particular, the first four population L-moments are:

$$\left\{\begin{array}{l} \lambda_1(X)=E\left(X_{[1:1]}\right) \\ \lambda_2(X)=(2)^{-1}E\left(X_{[2:2]}-X_{[1:2]}\right) \\ \lambda_3(X)=(3)^{-1}E\left[\left(X_{[3:3]}-X_{[2:3]}\right)-\left(X_{[2:3]}-X_{[1:3]}\right)\right] \\ \lambda_4(X)=(4)^{-1}E\left[\left(X_{[4:4]}-X_{[1:4]}\right)-3\left(X_{[3:4]}-X_{[2:4]}\right)\right]. \end{array}\right. \tag{2.3}$$

Knowing relation (2.1), they also satisfy the following equalities:

$$\begin{cases} \lambda_1(X) = \int_0^1 Q(u)\, du \\ \lambda_2(X) = \int_0^1 Q(u)(2u-1)\, du \\ \lambda_3(X) = \int_0^1 Q(u)(6u^2-6u+1)\, du \\ \lambda_4(X) = \int_0^1 Q(u)(20u^3-30u^2+12u-1)\, du. \end{cases} \tag{2.4}$$

Note that $\lambda_1(.)$, $\lambda_2(.)$, $\lambda_3(.)$ and $\lambda_4(.)$ are population measures of location, scale and shape, strictly analogous to the corresponding traditional central moments. The first L-moment is the mean of the population distribution and the second L-moment, defined in terms of a conceptual random sample of size 2, is a measure of the typical spread of the random variable X, being half the expected value of Gini's Mean Difference (see Gini (1912)), which has some desirable properties in finance when replacing the traditional variability measure (see Yitzhaki (2003)). In short, the third L-moment simply represents the difference between the upper tail and the lower tail, and hence measures the asymmetric shape of the population distribution from which the conceptual random sample has been drawn. Similarly, the fourth L-moment measures the *kurtosis* of the probability distribution function, and can be expressed as a (rescaled) difference between the typical spread in the tails and the typical spread in the center.

While Equation (2.1) is the classical L-functional's representation for the population L-moments, there exist several other representations for $\lambda_k(.)$ that prove to be useful for financial applications (see below). For example, using the definition of the Probability Weighted Moments, we can also express Equation (2.1) as a linear function of the Probability Weighted Moments, that is (see Hosking (1989)):

$$\lambda_k(X) = \sum_{r=1}^{k} p^*_{k-1,\,r-1}\, \beta_{r-1}(X), \tag{2.5}$$

with:

$$\beta_{r-1}(X) = r^{-1} E\left(X_{[r:r]}\right) = \int_0^1 Q(u)\, u^{r-1}\, du,$$

where $k \in IN^*$, $\beta_r(.)$ are the Probability Weighted Moments of order r, with $r = [2, ..., k]$, and $p^*_{k-1,r-1}$ is the $(r-1)$-*th* coefficient of the shifted Legendre polynomial of degree $(k-1)$ defined in Equation (2.1).

Using the definition and orthogonal property of the shifted Legendre polynomials, the k-*th* population L-moment can also be represented as the covariance between the random variable X and its distribution function denoted $F(.)$, that is (see Serfling and Xiao (2007)):

$$\lambda_k(X) = \begin{cases} E(X) & \text{if } k = 1 \\ Cov\left\{X, P^*_{k-1}[F(X)]\right\} & \text{if } k \neq 1, \end{cases} \tag{2.6}$$

with:

$$E\left\{P^*_k[F(X)]\right\} = \int_0^1 P^*_k(u)\, du = 0,$$

where $P^*_0(u) = 1$, with $k \in IN^*$, $P^*_k(.)$ the shifted orthogonal Legendre polynomial of degree k defined as previously, and $Cov(.,.)$ the covariance operator.

Since $E[F(X)] = 1/2$, relation (3.22) allows us to obtain alternative expressions for the first four population L-moments such as:

$$\begin{cases} \lambda_1(X) = E(X) \\ \lambda_2(X) = 2\,E\left\{[X - E(X)] \times [F(X) - E[F(X)]]\right\} \\ \lambda_3(X) = 6E\left\{[X - E(X)] \times [F(X) - E[F(X)]]^2\right\} \\ \lambda_4(X) = 20E\left\{[X - E(X)] \times [F(X) - E[F(X)]]^3\right\} - 3\,(2)^{-1}\,\lambda_2(X), \end{cases} \tag{2.7}$$

where $\lambda_1(.)$ and $\lambda_2(.)$ represent respectively the (L-)mean and the L-variance, and $\lambda_3(.)$ and $\lambda_4(.)$ correspond to the unscaled L-skewness and L-*kurtosis* of the population distribution $F(.)$.

The population L-moments presented in Equation (2.7) are defined for a given probability distribution, but, in practice, they are directly estimated with some uncertainty from a finite sample draw, corresponding to an unknown distribution. The following sub-section is devoted to a brief presentation of the (original) sample counterparts of the population

L-moments.

2.2.2 Sample Portfolio Return L-moments

Following Hosking (1989), sample L-moments can be estimated in a straightforward manner by estimating the empirical Probability Weighted Moments, denoted $\hat{\beta}_r(.)$, in conjunction with their L-statistic representations in (3.21).

Indeed, let $X_{[1:T]} \leq X_{[2:T]} \leq ... \leq X_{[T:T]}$ be again the order statistics of a random sample $\{X_t\}$ of size T, drawn from a continuous non-degenerated probability distribution $F(.)$ of a real-valued random variable X, with $t = [1, ..., T]$ and $T \in IN^*$. The r-*th* Probability Weighted Moment estimator is given, $\forall r \in IN^*$ and $1 < r < T$, by:

$$\widehat{\beta}_r(X) = (T)^{-1} \times \sum_{t=1}^{T} \left\{ \prod_{j=1}^{r} \left[\frac{(t-j)}{(T-j)} \right] X_{[t:T]} \right\}, \tag{2.8}$$

where $\widehat{\beta}_r(.)$ is the unbiased estimator of $\beta_r(.)$.

Thus, for a sample of size T, the k-*th* sample L-moment is defined as:

$$\widehat{\lambda}_k(X) = \sum_{r=1}^{k} p^*_{k-1,r-1} \widehat{\beta}_{r-1}(X), \tag{2.9}$$

where $\widehat{\lambda}_k(X)$ is the k-*th* sample L-moment corresponding to the population L-moment $\lambda_k(.)$ of a real-valued random variable X, $(r \times k \times T) \in (IN^*)^3$ with $r < k < T$, $p^*_{r-1,k-1}$ is the $(r-1)$-*th* coefficient of the shifted Legendre polynomial of degree $(k-1)$ defined as in Equation (2.1).

The L-moment estimator (2.9) is an unbiased L-statistic estimator of the population L-moment. In particular, in a time-series context, the first four sample L-moments from

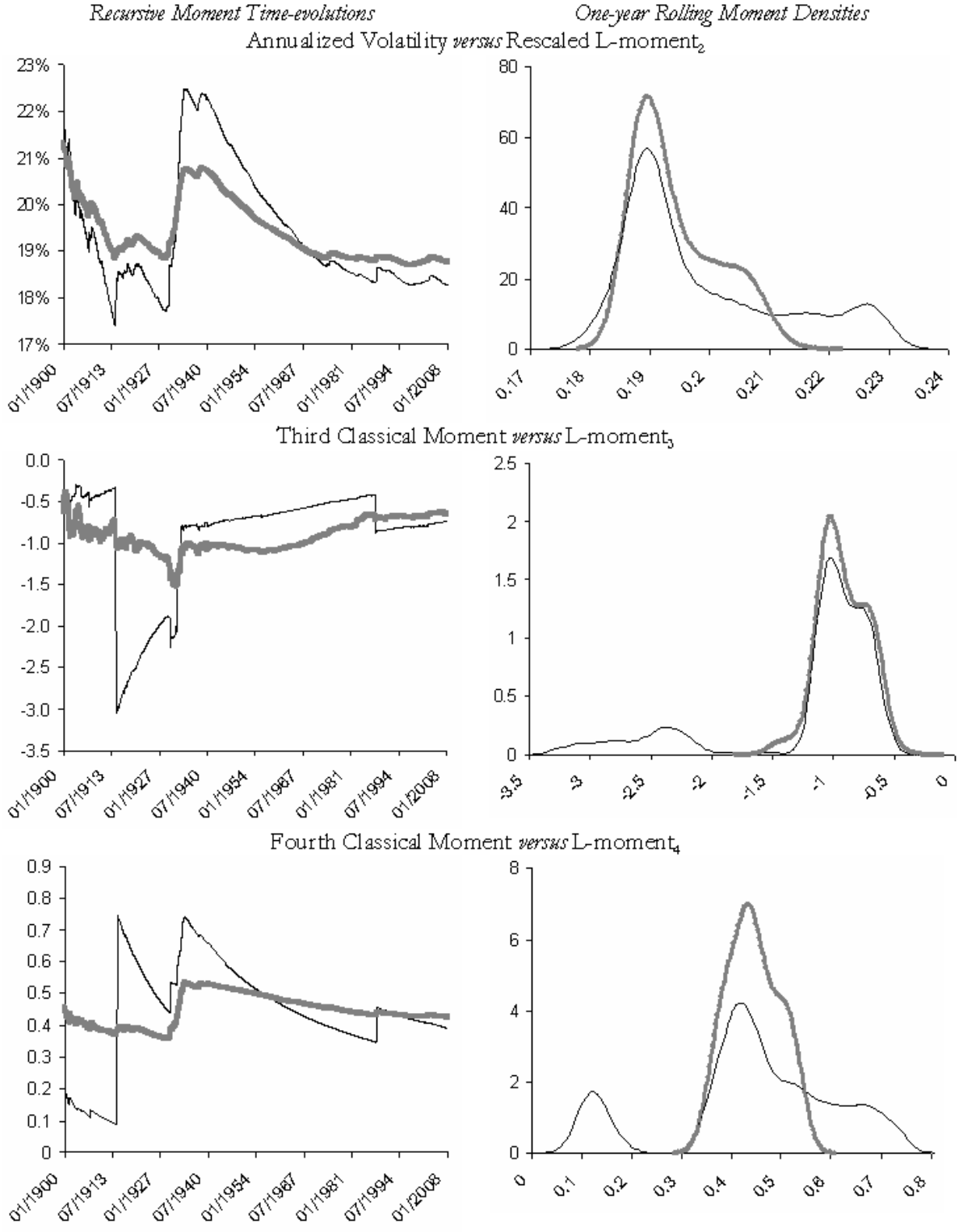
(2.9) are given by (with previous notations):

$$\begin{cases} \widehat{\lambda}_1(X) = & (T)^{-1} \sum_{t=1}^{T} X_{[t:T]} \\ \widehat{\lambda}_2(X) = & [T(T-1)]^{-1} \sum_{t=1}^{T} (2t-1-T)\ X_{[t:T]} \\ \widehat{\lambda}_3(X) = & [T(T-1)(T-2)]^{-1} \sum_{t=1}^{T} [6(t-1)\ \ t-6(t-1)\ \ T \\ & +(T-1)(T-2)]\, X_{[t:T]} \\ \widehat{\lambda}_4(X) = & [T(T-1)(T-2)(T-3)]^{-1} \\ & \times \sum_{t=1}^{T} [20(t-1)(t-2)(t-3) - 30(t-1)(t-2)(T-3) \\ & +12(t-1)(T-2)(T-3) - (T-1)(T-2)(T-3)]\, X_{[t:T]}. \end{cases} \tag{2.10}$$

Under the second central moment existence condition, the standard theory for U-statistics and L-statistics states that the (first) *k-th* sample L-moments are asymptotically jointly normally distributed, with a similar result for the vector of the sample L-moment ratios (see, for instance, Hoeffding (1948)).

Figure 2.1 illustrates the first (rescaled) L-moment estimations against the traditional sample moments, calculated on a long sample of one century of daily quotes of the Dow Jones Index. The set of figures on the left part (Figure 2.1) corresponds to time-evolutions of the second, third and fourth moments, recursively computed since the 1^{st} of January 1900, whilst the set of right figures represents the one-year rolling window first moment non-parametric empirical densities. As expected, it is clear from these figures that sample L-moments are far more stable than C-moment estimates. Moreover, densities of L-moments are more concentrated around a unique mode of L-moment values and exhibit fewer extreme values, indicating faster decreasing tails. This visually confirms that higher-order L-moments are less prone to the influence of outliers, and thus may be seen as more accurate.

Figure 2.1. Comparison of Classical Moments *versus* L-moments.



Source: *Bloomberg*, Daily Net Asset Values (01/01/1896-01/18/2008) in USD; computations by the authors. On the left side, the y-axis corresponds to the level (times 1.106 for the third and fourth moments) of the recursive classical moments (in black thin line); the corresponding recursive L-moments (in bold grey line) are rescaled in order to get the same means as their related Conventional moments. On the right side, all various moments are estimated using a one-year rolling window. The annual daily Lmoments are rescaled in order to have the same means as their related C-moments. The nonparametric densities are estimated with a Gaussian kernel using the cross-validation *criterion* (see Silverman (1986)). The probability density function of the classical moments corresponds to black thin lines whilst the one of the L-moments is in bold grey.

The main properties of population L-moments and their corresponding sample quantities now stated and their empirical robustness illustrated, we generalize in the next section, by using the shortage function approach of Luenberger (1995), the mean-variance efficient frontier in the first four L-moment space.

2.3 Portfolio Selection with Higher-order Moments

In the context of portfolio selection, the aim of the investors is to determine their asset allocation in order to maximize their utility function. We refer here to a general class of utility functions exhibiting a "mixed" risk aversion respecting the fourth-order stochastic dominance criterion (see Caballé and Pomansky (1996)), which alternate the signs of partial derivatives. In such a setting, we thus consider an exact (or accurate approximative) fourth-order Taylor expansion of a general utility function (see for details Jurzcenko and Maillet (2006a), and Garlappi and Skoulakis (2008)) with a strictly (monotone) increasing first derivative representative of the preference of non-satiable individuals, a strictly decreasing second derivative for risk-averse agents, a strictly increasing third derivative for prudent investors, and a strictly decreasing fourth derivative for temperate behaviors. More precisely (see Appendix 1 for a few illustrations with some usual utility functions), the expected utility of the random return on a portfolio p (denoted R_p) held by a rational investor, can be represented by an indirect utility function, denoted $V(.)$, successively, concave and increasing with the expected return - denoted $E(R_p)$, concave and decreasing with the variance - reading $\sigma^2(R_p)$, concave and increasing with the skewness - written $m^3(R_p)$, and, concave and decreasing with the *kurtosis*[4] - defined by $\kappa^4(R_p)$. Such an expected utility function can be written in a general form as:

$$E[U(R_p)] = V\left[E(R_p), \sigma^2(R_p), m^3(R_p), \kappa^4(R_p)\right], \tag{2.11}$$

with:

[4]Whilst, in general, skewness and *kurtosis* correspond to the standardised third and fourth centered moment, they are used here as the third and fourth centered moment.

$$V_1 = \frac{\partial V(.)}{\partial E(R_p)} > 0, V_2 = \frac{\partial V(.)}{\partial \sigma^2(R_p)} < 0, V_3 = \frac{\partial V(.)}{\partial m^3(R_p)} > 0 \text{ and } V_4 = \frac{\partial V(.)}{\partial \kappa^4(R_p)} < 0,$$

where $R_p = W/W_0 - 1$ is the (random) return on the portfolio p held by the investors, with W_0 their initial wealth (being equal to unity for the sake of simplicity), W their random final wealth, and $V(.)$ a general (non-)linear indirect utility function whose arguments are the first four conventional moments of returns on portfolio p.

The various first derivatives of such a general indirect utility function characterize at the same time both economic agent behavioural assumptions - his rational reaction to increases in downside risk, fear of ruin, will of self-protection and self-insurance (see Chiu (2005 and 2008), Crainich and Eeckhoudt (2008)), and a (simple) transformation of a density function of his return on wealth (see, for instance, Eeckhoudt, Gollier and Schneider (1995)). More precisely, the first derivative with respect to the expected return governs the so-called "greediness" of the investor, the second sensitivity represents his "risk aversion", whilst the third[5] and fourth terms characterize respectively the "prudence" (Kimball (1990) and Lajeri-Chaherli (2004)) and "temperance" (Kimball (1992 and 1993), Eeckhoudt, Gollier and Schneider (1995), Menezes and Wang (2005))[6].

From a theoretical point of view, the link between moments and preferences (Scott and Horwath (1980)) is still under question for at least four main reasons. First, moments are only single statistics that can only be imperfect summaries of the plain return distribution characteristics (see Romano and Siegel (1986))[7]. Depending on the exact transfer (preserving function) of probability weights, we can imagine all sorts of density distortions that basically break the rationale of the investor choice when comparing asset allocations (see

[5]Some interesting recent works, however, also show that the ratio $[U^{'''}(.)/U^{'}(.)]$ is also linked to a risk aversion characteristic of a rational agent, who makes an arbitrage between the first and the third moments (*Cf.* Crainich and Eeckhoudt (2008)).

[6]Lajeri-Chaherli (2004) extends the expansion to the order five, mentioning in reference the fifth-order risk apportionment "edginess", whilst Caballé and Pomansky (1996) refine even further the expansion to the *n-th* order, referring to the "risk aversion of order n", as an analogue to the traditional classical absolute risk aversion (see also Eeckhoudt and Schlesinger, (2006)). However, to our knowledge, no more precise label of utility characteristic yet exists for extensions to higher-order moments than the fifth. In the following, we nevertheless restrict our analysis to the four first moments, mainly for the sake of tractability, but also due to some questions regarding the existence of empirical counterparts of higher-order C-moments.

[7]The counterexample they mentioned being that the apparently left-skewed distribution of $x = \{-2, 1, 3\}$ with associated probabilities of $f(x) = \{.4, .5, .1\}$, has a null skewness.

Brockett and Kahane (1992) for some explicit examples). Secondly, first moments merely share the same information contained in the return series: the higher the order (and the power of the conventional moment), the more important the focus on the tail and extreme events *ceteris paribus*. Hence, some of them exhibit certain correlations *per* construction. In other words, redundant information[8] is present in the various comoments (see Galagedera and Maharaj (2008)). Thirdly, the explicit expressions of links between moments and preferences strongly depend upon the precise preference definition (see Haas (2007)) and on the performance of measures of higher-order moments (see Kim and White (2004), for various measures). Fourthly, in a more general prospect theory framework (Kahneman and Tversky (1979)), the rational investor may also further relax the linearity-in-probability property inherent in expected utility theory, by allowing the physical probabilities to be nonlinearly subjectively transformed into "decision weights" (see Kliger and Levy (2008)).

Nevertheless, from a more practical point of view, a wealth of literature (see Jondeau and Rockinger (2006), Jurczenko and Maillet (2006a), and Briec and Kerstens (2007), for a precise reference list on the subject) points out a realistic positive preference of investors for the highest right asymmetries and the lowest tail-fatnesses; we will take this common sense *hypothesis* as granted in the following, where we simply try to extend the two-moment Markowitz' analysis in an expected utility framework, only considering higher moments with better properties.

Since we will also only consider "mixed" utility functions, the signs of the sensitivities V_n (partial derivatives), for $n = [1, ..., 4]$, will alternate. Furthermore, in a portfolio context, one could intuitively expect that the investor cares more about a (positive) expected return than about other characteristics[9], and, as a result, that the sensitivities decrease

[8]It is straightforward to show, for instance, that some terms in the conventional *cokurtosis* matrix also appear in both covariance and coskewness matrices (see among others Jurczenko and Maillet (2006a), on notations of higher-order co-moments).

[9]It is difficult to believe that most rational investors care more about higher-order moments than about the expected return in their asset allocation decisions (which can be the case in a lottery for instance, where the potential big prize may entail a large skewness, allowing them to forget in some sense the likely negative profit of the game). This is the reason why in the following, we further restrict the search for optimal portfolios in regions where the expected returns are positive, and only select, in some representations, portfolios where impacts on the utility functions of moments are ranked according to their order (see Section 4).

with the order of related moments.

In such a framework, the agent's portfolio general problem can be stated as (with previous notations):

$$\begin{array}{rcl} \underset{\mathbf{w}_p^{'}}{\text{Max}} \left\{ E\left[U\left(R_p \right) \right] \right\} & = & \underset{\mathbf{w}_p^{'}}{\text{Max}} \left\{ V\left[E\left(R_p \right), \sigma^2\left(R_p \right), m^3\left(R_p \right), \kappa^4\left(R_p \right) \right] \right\} \\ s.t. & : & \mathbf{w}_p^{'} \mathbf{1}_N = 1, \end{array} \tag{2.12}$$

where $\mathbf{1}_N$ is the $(N \times 1)$ unit vector and $V(.)$ a general non-explicited (non-)linear function depending on the four first C-moments of returns on portfolio p that we explicit hereafter.

2.3.1 Higher-order C-comoments of Portfolio Returns

Actually, the mean, variance, skewness and *kurtosis* of portfolio p returns used in Equation (11) are given by, with $(i,\, j,\, k,\, l) = [1, ..., N]^4$ (and with previous notations):

$$\left\{ \begin{array}{l} E\left(R_p \right) = \sum_{i=1}^{N} w_{pi}\, E\left(R_i \right) \\ \sigma^2\left(R_p \right) = E\left\{ \left[R_p - E\left(R_p \right) \right]^2 \right\} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{pi} w_{pj}\, \sigma_{ij} \\ m^3\left(R_p \right) = E\left\{ \left[R_p - E\left(R_p \right) \right]^3 \right\} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} w_{pi} w_{pj} w_{pk} m_{ijk} \\ \kappa^4\left(R_p \right) = E\left\{ \left[R_p - E\left(R_p \right) \right]^4 \right\} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} w_{pi} w_{pj} w_{pk} w_{pl} \kappa_{ijkl}, \end{array} \right. \tag{2.13}$$

with:

$$\left\{ \begin{array}{l} \sigma_{ij} = E\left\{ \left[R_i - E\left(R_i \right) \right] \left[R_j - E\left(R_j \right) \right] \right\} \\ m_{ijk} = E\left\{ \left[R_i - E\left(R_i \right) \right] \left[R_j - E\left(R_j \right) \right] \left[R_k - E\left(R_k \right) \right] \right\} \\ \kappa_{ijkl} = E\left\{ \left[R_i - E\left(R_i \right) \right] \left[R_j - E\left(R_j \right) \right] \left[R_k - E\left(R_k \right) \right] \left[R_l - E\left(R_l \right) \right] \right\}, \end{array} \right.$$

where w_{pi}, R_i, σ_{ij}, m_{ijk} and κ_{ijkl} represent, respectively, the weight of the asset i in portfolio p, the return on the asset i, the covariance between the returns on assets i and j, the coskewness between the returns on assets i, j and k, and the *cokurtosis* between the

returns on assets i, j, k and l.

These various C-moments of portfolio returns were previously written in a matrix format (see Diacogiannis (1994), Athayde and Florès (2002, 2003, 2004 and 2006), Harvey, Liechty, Liechty, Mueller (2002), Prakash, Chang and Pactwa (2003), Jondeau and Rockinger (2003a, 2003b and 2006) and Jurczenko, Maillet and Merlin, (2006)) defined as such (with previous notations):

$$\begin{cases} E\left(R_p\right) = \mathbf{w}_p^{'}\,\mathbf{E} \\ \sigma^2\left(R_p\right) = \mathbf{w}_p^{'}\,\mathbf{\Omega}\,\mathbf{w}_p \\ m^3\left(R_p\right) = \mathbf{w}_p^{'} \times \mathbf{\Sigma} \times \ \left(\mathbf{w}_p \otimes \mathbf{w}_p\right) \\ \kappa^4\left(R_p\right) = \mathbf{w}_p^{'} \times \mathbf{\Gamma} \ \times \ \left(\mathbf{w}_p \otimes \mathbf{w}_p \otimes \mathbf{w}_p\right), \end{cases} \tag{2.14}$$

where $\mathbf{w}_p$ is the weight vector of assets in p, $\mathbf{E}$ is the $(N \times 1)$ vector of expected returns, $\mathbf{\Omega}$ is the $(N \times N)$ matrix of covariance, $\mathbf{\Sigma}$ is the $(N \times N^2)$ global matrix of coskewness, and $\mathbf{\Gamma}$ is the $(N \times N^3)$ global matrix of *cokurtosis* between all risky security returns, and the sign $\otimes$ standing for the symbol of the Kronecker product[10].

In this representation, matrices $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$ are built using the following scheme:

$$\begin{cases} \underset{(N\times N^2)}{\mathbf{\Sigma}} = \left(\mathbf{\Sigma}_1 \ \mathbf{\Sigma}_2 \ \cdots \ \mathbf{\Sigma}_N\right) \\ \underset{(N\times N^3)}{\mathbf{\Gamma}} = \left(\mathbf{\Gamma}_{11} \ \mathbf{\Gamma}_{12} \ \cdots \ \mathbf{\Gamma}_{1N} \mid \mathbf{\Gamma}_{21} \ \mathbf{\Gamma}_{22} \ \cdots \ \mathbf{\Gamma}_{2N} \mid ... \mid \mathbf{\Gamma}_{N1} \ \mathbf{\Gamma}_{N2} \ \cdots \ \mathbf{\Gamma}_{NN}\right) \end{cases} \tag{2.15}$$

[10]If $\mathbf{A}$ is a $(M \times P)$ matrix and $\mathbf{B}$ a $(N \times Q)$ matrix, the $(MN \times PQ)$ matrix $(\mathbf{A} \otimes \mathbf{B})$ is called the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, and is defined as such:

$$\underset{(MN\times PQ)}{\mathbf{A} \otimes \mathbf{B}} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1P}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2P}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MP}\mathbf{B} \end{pmatrix},$$

where:

$$\underset{(N\times Q)}{a_{np}\mathbf{B}} = \begin{pmatrix} a_{np}b_{11} & a_{np}b_{12} & \cdots & a_{np}b_{1Q} \\ a_{np}b_{21} & a_{np}b_{22} & \cdots & a_{np}b_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{np}b_{N1} & a_{np}b_{N2} & \cdots & a_{np}b_{NQ} \end{pmatrix},$$

with a_{mp} and b_{nq} are the elements of matrices A and B, and $(m, n, p, q) = [1, ..., M] \times [1, ..., N] \times [1, ..., P] \times [1, ..., Q] \subset IN^{*4}$.

where $\mathbf{\Sigma}_k$ and $\mathbf{\Gamma}_{kl}$ are the $(N \times N)$ associated sub-matrices of $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$, with single elements $(s_{ijk})_{(i,j)=[1,..,N]^2}$ and $(\kappa_{ijkl})_{(i,j)=[1,..,N]^2}$, for any given coupled $(k \times l) = (IN^*)^2$.

We now propose herein a strictly equivalent notation for defining C-moments. Let us first start by defining the *n-th* recursive convolution matrix operator of a function $H(.)$, denoted *per* convention $H^{(\bigcirc n)}(\mathbf{w}_p)$, for $n \in IN^*$, such as:

$$H^{(\bigcirc n)}(.) \equiv \underbrace{H\{H[...H(.)]\}}_{n \text{ operations}}, \tag{2.16}$$

and with for $n = 0$, *per* definition, $H^{(\bigcirc 0)}(.) = Id(.)$ the identity function.

Secondly, we can define the following recurrent relation, applied to weight $\mathbf{w}_p$, with $n \in IN$ (and with previous notations):

$$H^{(\bigcirc n)}(\mathbf{w}_p) = \mathbf{Vec}\left[H^{[\bigcirc(n-1)]}(\mathbf{w}_p) \times \mathbf{w}_p^{'}\right] = H^{[\bigcirc(n-1)]}(\mathbf{w}_p) \otimes \mathbf{w}_p, \tag{2.17}$$

where the function $H(.)$ is defined such as $H(\mathbf{W}_p) = \mathbf{Vec}\left(\mathbf{W}_p \times \mathbf{w}_p^{'}\right)$, with $\mathbf{W}_p$ being a vector of transformed weights $\mathbf{w}_p$ and $\mathbf{Vec}(.)$ the operator that reshapes a $(N \times M)$ matrix in a $(NM \times 1)$ vector, with $(N, M) = IN^{*2}$.

Thirdly, define the (repeated) Hadamard product of returns on any set of n assets of the portfolio p under studies, such that (*per* convention) we have, with $n \in [2, ..., 4]$:

$$\underset{\substack{q=1 \\ (T\times 1)}}{\overset{n}{\bigodot}} \check{\mathbf{R}}\left(a_{[q]}\right) \equiv \underbrace{\check{\mathbf{R}}\left(a_{[1]}\right) \odot ... \odot \check{\mathbf{R}}\left(a_{[n]}\right)}_{n \text{ terms}}, \tag{2.18}$$

where the sign $\odot$ stands for the symbol of the (simple) Hadamard product[11], and with $\check{\mathbf{R}}(q)$ the *q-th* column of $\check{\mathbf{R}}$, $\check{\mathbf{R}} = \mathbf{R} - (\mathbf{E} \times \mathbf{1}_T')'$ being the $(T \times N)$ matrix of centered returns, $\mathbf{R}$ the $(T \times N)$ matrix of returns on the N assets, $\mathbf{1}_T$ the $(T \times 1)$ unit vector and the $a_{[q]}$ (with $q \in [1, ..., n] \subset [1, ..., N]$) being the ranks (column number) of the assets in the matrix of excess returns $\check{\mathbf{R}}$ (taken in any order), that we want to compute the related higher-order comoment, and which identify the location of a specific element in global matrices of higher-order comoments of individual stock returns.

With the two previous definitions and the recurrent relation, we are now able to define any (scalar) C-moment[12] of order n, denoted $m^n(R_p)$, as well as any related global (higher-order) comoment $(N \times N^{n-1})$ matrix $\mathbf{M}^n$, with elements[13] $\mathbf{M}^n_{(i,j)=(N \times N^{n-1})}$, with $j = \sum_{q=1}^{n-1} \left(a_{[q]} N^{n-1-q}\right)$, being such that, with $n \in [2, ..., 4]$ (and with previous notations):

[11]The $(N \times M)$ Hadamard product matrix $(\mathbf{A} \odot \mathbf{B})$ of two similar $(N \times M)$ matrices $\mathbf{A}$ and $\mathbf{B}$, is defined as such:

$$\underset{(1\times 1)}{(\mathbf{A} \odot \mathbf{B})_{(n,m)}} = \mathbf{A}_{(n,m)} \times \mathbf{B}_{(n,m)}$$
$$\Longleftrightarrow$$
$$\underset{(N\times M)}{\mathbf{A} \odot \mathbf{B}} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1M}b_{1M} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2M}b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}b_{N1} & a_{N2}b_{N2} & \cdots & a_{NM}b_{NM} \end{pmatrix},$$

where a_{nm} and b_{nm} are elements of A and B, with $(n, m) = [1, ..., N] \times [1, ..., M] \subset IN^{*2}$.

[12]This notation can furthermore be extended to higher higher-order L-comoments with no difficulty. For instance, the fifth-order Linear moment and comoment read (with the same notations):

$$\begin{cases} m^5(R_p) = \mathbf{w}_p' \times \mathbf{M}^5 \times H^{(\bigcirc 5)}(\mathbf{w}_p) \\ m_{ijklm} = \mathbf{M}^5{}_{[i,(j-1)N^3+(k-1)N^2+(l-1)N+m]} = T^{-1} \times \mathbf{1}_T' \times \left[\check{\mathbf{R}}(i) \odot \check{\mathbf{R}}(j) \odot \check{\mathbf{R}}(k) \odot \check{\mathbf{R}}(l) \odot \check{\mathbf{R}}(m)\right]. \end{cases}$$

Since L-moments always exist, this expression may interestingly lead to further future refinements.

[13]As pointed out by Jondeau, Poon and Rockinger (2007), many elements are the same in the matrices $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$: only $N(N+1)(N+2)/6$ out of N^3 for $\mathbf{\Sigma}$ and $N(N+1)(N+2)(N+3)/24$ out of N^4 for $\mathbf{\Gamma}$ are different. Imposing the $a_{[q]}$, $q \in [1, ..., n] \subset [1, ..., N]$, the ranks of the assets (that we want to compute a specific higher-order comoment) to be ordered as in the matrix $\check{\mathbf{R}}$ (and not free as in the above notation), allows us to only provide the distinct elements of the matrix $\mathbf{M}^n$. For illustration purpose, we note that for the 162 stocks used in the following empirical application (see below), the coskewness and *cokurtosis* matrices contain more than 3 millions and 650 millions redundant terms. Not computing these terms and weighting the distinct ones according to the number of their repetitions, leads us to divide by ten or so the computation time of matrix $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$. Moreover, the parsimonious new approach permits us to handle large-scale portfolio problems more easily.

$$\left\{\begin{array}{lcl} \underset{(1\times1)}{m^n\left(R_p\right)} & = & \mathbf{w}_p^{'}\times\mathbf{M}^n\times H^{[\bigcirc(n-2)]}\left(\mathbf{w}_p\right) \\ \underset{(1\times1)}{m_{a_{[1]}\ldots a_{[n]}}} & = & \mathbf{M}^n_{\left[a_{[1]},\sum\limits_{q=1}^{n-1}\left(a_{[q+1]}-1\right)\times N^{n-1-q}\right]}=T^{-1}\times\mathbf{1}_T^{'}\times\left[\bigodot_{q=1}^{n}\check{\mathbf{R}}\left(a_{[q]}\right)\right]. \end{array}\right. \tag{2.19}$$

Using this generic writing, the system of C-moments in equation (2.14) then can simply be expressed as such (with previous notations):

$$\left\{\begin{array}{l} E\left(R_p\right)=\mathbf{w}_p^{'}\mathbf{E}\times1 \\ \sigma^2\left(R_p\right)=\mathbf{w}_p^{'}\times\mathbf{\Omega}_{\mathbf{w}_p}=\mathbf{w}_p^{'}\times\left(\mathbf{\Omega}\times\mathbf{w}_p\right)\equiv\mathbf{w}_p^{'}\times\left[\mathbf{\Omega}\times H^{(\bigcirc0)}\left(\mathbf{w}_p\right)\right] \\ m^3\left(R_p\right)=\mathbf{w}_p^{'}\times\mathbf{\Sigma}_{\mathbf{w}_p}=\mathbf{w}_p^{'}\times\left[\mathbf{\Sigma}\times\left[H\left(\mathbf{w}_p\right)\right]\right]\equiv\mathbf{w}_p^{'}\times\left[\mathbf{\Sigma}\times H^{(\bigcirc1)}\left(\mathbf{w}_p\right)\right] \\ \kappa^4\left(R_p\right)=\mathbf{w}_p^{'}\times\mathbf{\Gamma}_{\mathbf{w}_p}=\mathbf{w}_p^{'}\times\left\{\mathbf{\Gamma}\times H\left[H\left(\mathbf{w}_p\right)\right]\right\}\equiv\mathbf{w}_p^{'}\times\left[\mathbf{\Gamma}\times H^{(\bigcirc2)}\left(\mathbf{w}_p\right)\right], \end{array}\right. \tag{2.20}$$

where $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$ are (still) the global $(N\times N^2)$ coskewness and $(N\times N^3)$ *cokurtosis* matrices (strictly equivalent to their previous tensor forms), but with each element being expressed in the new notation such as, $\forall(i,\,j,\,k,\,l)=[1,...,N]^4$ (with previous notations):

$$\left\{\begin{array}{lcl} \underset{(1\times1)}{\sigma_{ij}} & = & \mathbf{\Omega}_{[i,j]}=T^{-1}\times\mathbf{1}_T^{'}\times\left[\check{\mathbf{R}}\left(i\right)\odot\check{\mathbf{R}}\left(j\right)\right] \\ \underset{(1\times1)}{m_{ijk}} & = & \mathbf{\Sigma}_{[i,(j-1)N+k]}=T^{-1}\times\mathbf{1}_T^{'}\times\left[\check{\mathbf{R}}\left(i\right)\odot\check{\mathbf{R}}\left(j\right)\odot\check{\mathbf{R}}\left(k\right)\right] \\ \underset{(1\times1)}{\kappa_{ijkl}} & = & \mathbf{\Gamma}_{[i,(j-1)N^2+(k-1)N+l]}=T^{-1}\times\mathbf{1}_T^{'}\times\left[\check{\mathbf{R}}\left(i\right)\odot\check{\mathbf{R}}\left(j\right)\odot\check{\mathbf{R}}\left(k\right)\odot\check{\mathbf{R}}\left(l\right)\right]. \end{array}\right.$$

The previous new compact notation exhibits some advantages compared to the traditional one (Athayde and Flôres (2002)), which essentially relied on a tensor notation of coskewness and *cokurtosis* of asset returns on a portfolio, its global skewness and *kurtosis*. This new notation, more "computational-oriented", is strictly equivalent to the previous one (since the Hadamard product terms are all included in the Kronecker matrices of weights), but first can be generalized in a more compact form in a recursive manner from the first moment to the *n-th* higher than the fourth, and, furthermore, gives a direct expression of all elements of the skewness and *kurtosis* matrices; secondly, it still allows us

to disentangle the weight and the asset return impacts on higher-order moments; thirdly, it explicits the links between higher-order comoments and fourthly, it uses only traditional simple (low-level) pre-programmed operators (for building matrices of coskewness and *cokurtosis*) and thus appears to lead to a substantial overall gain in terms of execution time[14].

The traditional multi-moment asset allocation setting now revisited, we shall adapt it hereafter to the analogues of the above C-moments of returns on any portfolio p, in the robust framework of Linear moments.

2.3.2 Higher-order L-comoments of Portfolio Returns

Let us also recall that $\mathbf{R}$, $\mathbf{E}$ and $\check{\mathbf{R}}$ denote respectively the $(T \times N)$, $(N \times 1)$ and $(T \times N)$ vectors of effective returns, expected returns and centered realized returns on the N risky assets. In the context of robust L-comoment computations, the expectation (denoted $E\left(R_{\mathbf{w}_p}\right)$, with $R_{\mathbf{w}_p}$ being a random variable) of the $(N \times 1)$ vector of observed returns, $\mathbf{R}_{\mathbf{w}_p}$, on the portfolio defined by its weight $\mathbf{w}_p$, the matrices $\mathbf{\Omega}^{(L)}_{\mathbf{w}_p}$, $\mathbf{\Sigma}^{(L)}_{\mathbf{w}_p}$ and $\mathbf{\Gamma}^{(L)}_{\mathbf{w}_p}$, representing respectively the $(N \times 1)$ vectors of the L-covariance, L-coskewness and L-*cokurtosis*[15] of the security returns with the returns on the portfolio p, can be defined[16]

[14]When we empirically double-checked the strict equivalence of the two alternative notations of higher-order comoments (*i.e.* the traditional Kronecker *versus* the new recursive Hadamard forms), it appears that the new one leads to a (limited) reduction of the execution time (by 7% or so), representing, however, several hours of spared computation time in large-scale portfolio selection applications. Moreover, both previous notations using C-comoments include a lot of redundant information in global matrices of coskewness and *cokurtosis* (see previous Footnote). Only computing the distinct elements may represent a 90% economy of execution time in a large-scale problem (see below).

[15]Note that the dimension of the second C-comoment matrix $\mathbf{\Omega}$ and those of the second L-comoment matrix $\mathbf{\Omega}^{(L)}_{\mathbf{w}_p}$, are different; the same is true for higher-order comoments (*i.e.* $\mathbf{\Sigma}_{\mathbf{w}_p}$ and $\mathbf{\Sigma}^{(L)}_{\mathbf{w}_p}$, $\mathbf{\Gamma}_{\mathbf{w}_p}$ and $\mathbf{\Gamma}^{(L)}_{\mathbf{w}_p}$).

[16] The first L-moment strictly corresponds to the arithmetic mean return. Since we propose robust statistics to assess portfolio return peculiarities, it was natural to wonder if the use of the alternative measures of expected performance could have been more appropriate in a portfolio choice context. Some authors have shown that a bias could arise when using an arithmetic mean; a geometric mean is certainly more accurate when estimating a long term expected return. But with the one-week horizon used in our application, this bias is not relevant (see Hughson, Stutzer and Yung (2006)). Other authors prefer to use a robust statistic for a location parameter (such as the median) instead of the classical mean when estimating expected returns (see McCulloch (2003)). Two reasons motivate us to stay here in the classical paradigm. First, the median (or the first Trimmed L-moment) neglects the impact of (some of) the extreme returns on the performance; considering the median could thus in some cases blur the investor perception. Secondly, we also computed Four-moment optimal portfolios using the median, but we did

as (with previous notations):

$$\left\{\begin{array}{ll} \underset{(1\times 1)}{E\left(R_{\mathbf{w}_p}\right)} & = \mathbf{w}_p^{'}\mathbf{E} \\ \underset{(N\times 1)}{\mathbf{\Omega}_{\mathbf{w}_p}^{(L)}} & = 2\,E\left\{\check{\mathbf{R}}\times\left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right)-E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}\right\} \\ \underset{(N\times 1)}{\mathbf{\Sigma}_{\mathbf{w}_p}^{(L)}} & = 6\,E\left\{\check{\mathbf{R}}\times\left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right)-E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}^2\right\} \\ \underset{(N\times 1)}{\mathbf{\Gamma}_{\mathbf{w}_p}^{(L)}} & = 20\,E\left\{\check{\mathbf{R}}\times\left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right)-E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}^3\right\}, \end{array}\right. \tag{2.21}$$

where $F(.)$ is the distribution of the random variable $R_{\mathbf{w}_p}$.

Using the covariance representation of L-moments defined in Equation (3.22) and the bilinear property of the covariance operator (see, for instance, Yitzhaki (2003)), the various population L-(co)moments of the returns on any attainable portfolio are respectively given by (with previous notations):

$$\left\{\begin{array}{l} \lambda_1\left(R_{\mathbf{w}_p}\right)=E\left(R_{\mathbf{w}_p}\right)=\sum_{i=1}^{N} w_i E\left(R_i\right) \\ \lambda_2\left(R_{\mathbf{w}_p}\right)=2Cov[R_{\mathbf{w}_p},F\left(R_{\mathbf{w}_p}\right)]=\sum_{i=1}^{N} w_i\ \lambda\text{-}Cov\left(R_i,R_{\mathbf{w}_p}\right) \\ \lambda_3\left(R_{\mathbf{w}_p}\right)=6Cov\{R_{\mathbf{w}_p},\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\}^2\}=\sum_{i=1}^{N} w_i\ \lambda\text{-}Cos\left(R_i,R_{\mathbf{w}_p}\right) \\ \lambda_4\left(R_{\mathbf{w}_p}\right)=Cov\left\{R_{\mathbf{w}_p},20\left\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\right\}^3-3\left\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\right\}\right\} \\ \qquad\quad =\sum_{i=1}^{N} w_i\left[\lambda\text{-}Cokurt\left(R_i,R_{\mathbf{w}_p}\right)-3\left(2\right)^{-1}\lambda\text{-}Cov\left(R_i,R_{\mathbf{w}_p}\right)\right], \end{array}\right. \tag{2.22}$$

and, for any asset i, with $i\in[1,...,N]$:

$$\left\{\begin{array}{l} \lambda\text{-}Cov\left(R_i,R_{\mathbf{w}_p}\right)=2\,E\left\{\left[R_i-E\left(R_i\right)\right]\times\left\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\right\}\right\} \\ \lambda\text{-}Cos\left(R_i,R_{\mathbf{w}_p}\right)=6\,E\left\{\left[R_i-E\left(R_i\right)\right]\times\left\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\right\}^2\right\} \\ \lambda\text{-}Cokurt\left(R_i,R_{\mathbf{w}_p}\right)=20\,E\left\{\left[R_i-E\left(R_i\right)\right]\times\left\{F\left(R_{\mathbf{w}_p}\right)-E\left[F\left(R_{\mathbf{w}_p}\right)\right]\right\}^3\right\} \\ E\left[F\left(R_{\mathbf{w}_p}\right)\right]=1/2, \end{array}\right.$$

where λ-$Cov(.)$, λ-$Cos(.)$ and λ-$Cokurt(.)$ correspond respectively to the L-covariance, L-

not find any clear difference between the two approaches (same overall conclusions in Section 4 apply; see also Footnote 21).

coskewness and L-*cokurtosis* between any asset i return and the portfolio p return defined by its holdings $\mathbf{w}_p$.

The L-covariance, L-coskewness and L-*cokurtosis* of portfolio returns used in Equation (2.21) can be written in a compact matrix format[17] as such, for $n \geq 2$ (with previous notations):

$$\underset{(N\times 1)}{\mathbf{M}_{\mathbf{w}_p}^{(L)\ n}} = p^*_{n-1,\,n-1}\ T^{-1}\ \left\{\check{\mathbf{R}} \odot \left\{\bigodot_{q=1}^{n-1}\left\{q^0 \times \left[F\left(\mathbf{R}_{\mathbf{w}_p}\right) \times \mathbf{1}_N^{'} - 1/2\right]\right\}\right\}\right\}^{'} \times \mathbf{1}_T, \quad (2.23)$$

with $\mathbf{M}_{\mathbf{w}_p}^{(L)\ 2} = \mathbf{\Omega}_{\mathbf{w}_p}^{(L)}$, $\mathbf{M}_{\mathbf{w}_p}^{(L)\ 3} = \mathbf{\Sigma}_{\mathbf{w}_p}^{(L)}$, $\mathbf{M}_{\mathbf{w}_p}^{(L)\ 4} = \mathbf{\Gamma}_{\mathbf{w}_p}^{(L)}$ and $p^*_{n-1,\,n-1} = [2\,(n-1)]!\ [(n-1)!]^{-2}$ a factor being equal to the (n-1)-*th* (highest) coefficient of the shifted orthogonal Legendre polynomial $P^*_{n-1}(.)$ of degree $n-1$, as previously defined in equation (2.1).

The L-moments, written in a generic manner such as (with previous notations):

$$\begin{cases} \lambda_1\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{E} \\ \lambda_2\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{\Omega}_{\mathbf{w}_p}^{(L)} \\ \lambda_3\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{\Sigma}_{\mathbf{w}_p}^{(L)} \\ \lambda_4\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{\Gamma}_{\mathbf{w}_p}^{(L)} - 3\ (2)^{-1}\,\mathbf{w}_p^{'}\,\mathbf{\Omega}_{\mathbf{w}_p}^{(L)}, \end{cases} \quad (2.24)$$

These L-moments can be reformulated in a (even) more compact manner[18] reading, for $n \geq 2$:

$$\lambda_n\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{Mc}_{\mathbf{w}_p}^{(L)\ n}, \quad (2.25)$$

with:

$$\mathbf{Mc}_{\mathbf{w}_p}^{(L)\ n} = T^{-1}\left\{\check{\mathbf{R}} \odot \left\{\left\{P^*_{n-1}\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right] - E\left\{P^*_{n-1}\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}\right\} \times \mathbf{1}_N^{'}\right\}^{'} \times \mathbf{1}_T,$$

[17] Compared to higher-order C-comoment writings, we note here that L-comoment analogues do not contain any redundant element.

[18] The use of higher-order L-comoments, computed with this compact writing instead of the previous related C-moment one, leads empirically to a drastic reduction in the execution time (divided by about four or so) in our general Goal Attainement problem (see below), whilst computing the first L-moments instead of the first C-moments of portfolio returns is approximately 40% faster.

where $\mathbf{Mc}_{\mathbf{w}_p}^{(L)\,2} = \mathbf{\Omega}_{\mathbf{w}_p}^{(L)}$, $\mathbf{Mc}_{\mathbf{w}_p}^{(L)\,3} = \mathbf{\Sigma}_{\mathbf{w}_p}^{(L)}$, and $\mathbf{Mc}_{\mathbf{w}_p}^{(L)\,4} = \mathbf{\Gamma}_{\mathbf{w}_p}^{(L)} - 3\,(2)^{-1}\,\mathbf{\Omega}_{\mathbf{w}_p}^{(L)}$.

Since we now have the complete characterization of all L-moments of portfolio returns, we can then define hereafter more precisely the set of efficient portfolios.

2.3.3 Higher-order L-moments and the Efficient Frontier Definition

We now consider the problem of an investor selecting a portfolio from N risky assets (with $N \geq 4$) in the four L-moment framework. We assume that the investor does not have access to a riskless asset, and that the portfolio weights sum to one. In addition, we impose[19] a no short-sale portfolio constraint.

Any portfolio p is here entirely defined by $\mathbf{w}_p \in IR^N$, the vector of weights of assets, and the set of the attainable portfolios $\mathfrak{A}$ can then be expressed as follows:

$$\mathfrak{A} = \left\{ \mathbf{w} \in IR^N : \mathbf{w}^{'}\mathbf{1} = 1 \text{ and } \mathbf{w} \geq \mathbf{0} \right\}, \tag{2.26}$$

where $\mathbf{w}^{'}$ is the $(1 \times N)$ transposed vector of the investor's holdings in the various risky assets, $\mathbf{1}$ is the $(N \times 1)$ unitary vector and $\mathbf{0}$ is the $(N \times 1)$ null vector

As in Markowitz (1952), the definition of moments of portfolio's returns indeed leads to the disposal representation of the set of the feasible portfolios, denoted $\mathfrak{F}$, in the extended four L-moment space[20] (see Briec, Kerstens and Lesourd (2004), and Briec, Kerstens and Jokung (2007)), reading:

$$\mathfrak{F} = \{\lambda_{\mathbf{w}} : \mathbf{w} \in \mathfrak{A}\} + [(-IR_+) \times IR_+ \times (-IR_+) \times IR_+], \tag{2.27}$$

where $\lambda_{\mathbf{w}}$ is the (4×1) vector of the first four L-moments[21] of the portfolio return $R_{\mathbf{w}}$,

[19] This assumption can however be relaxed to some extent with no loss of generality (see Footnote 28).

[20] For the sake of simplicity, for each point of the efficient frontier, we choose to consider "portfolios" even if we should only speak about "classes of equivalence induced by these portfolios" (see end of Appendix 2).

[21] As mentioned earlier, we make use here of the traditional L-moments instead of other variants (such as the Trimmed L-moments for instance). Despite the fair argument by Darolles, Gouriéroux and Jasiak (2008), highlighting that for large samples the "...Trimmed L-moments of order 1 bridge the mean and

i.e.:

$$\lambda_{\mathbf{w}} = \left[\lambda_1\left(R_{\mathbf{w}}\right);\ \lambda_2\left(R_{\mathbf{w}}\right);\ \lambda_3\left(R_{\mathbf{w}}\right);\ \lambda_4\left(R_{\mathbf{w}}\right)\right]'.$$

This disposal representation is necessary here to ensure the convexity of the feasible portfolio set in the four L-moment space (see Briec and Kerstens, (2007), and Zhang (2008) for an illustration on consequences of the non-convexity of the set of portfolios on the choice of optimal ones).

We define a strict (generic) order relation, denoted by $\succ$, on IR^4, that is for any $(\lambda, \widetilde{\lambda}) \in (IR^4)^2$:

$$\lambda \succeq \widetilde{\lambda} \iff [\lambda_1 \geq \widetilde{\lambda}_1,\ \lambda_2 \leq \widetilde{\lambda}_2,\ \lambda_3 \geq \widetilde{\lambda}_3,\ \lambda_4 \leq \widetilde{\lambda}_4], \tag{2.28}$$

altogether with a strict relation, denoted by $\succ$, as such:

$$\lambda \succ \widetilde{\lambda} \iff [\lambda_1 > \widetilde{\lambda}_1,\ \lambda_2 < \widetilde{\lambda}_2,\ \lambda_3 > \widetilde{\lambda}_3,\ \lambda_4 < \widetilde{\lambda}_4]. \tag{2.29}$$

The four L-moment weakly efficient frontier $\mathfrak{L}$ is then defined as follows:

$$\mathfrak{L} = \left\{\lambda_{\mathbf{w}} \in \mathfrak{F} :\ \forall\ \tilde{\lambda} \in IR^4,\ \tilde{\lambda} \succ \lambda_{\mathbf{w}} \Rightarrow \tilde{\lambda} \notin \mathfrak{F}\right\}. \tag{2.30}$$

whilst the four L-moment strong efficient frontier $\mathfrak{M}$ is defined as follows:

the median", we choose not to use them in a portfolio choice context for three main reasons. As a first reason, it is clear that Trimmed or Quantile L-moments provide more accurate estimates of the underlying distribution characteristics when focusing on extremes in a risk estimation exercise for instance; it is far from obvious, however, that the influence of large deviations should be too much reduced in a portfolio choice framework, since extremes should have - in a sense - some influence on the first and the second moments of returns. Deleting some really "bad" returns of a hedge fund record for instance and computing the related first Trimmed L-moment will probably result in an upward bias of the future anticipated performance of the fund. Secondly, if there exist more robust alternatives to the sensitive mean operator for the location parameter of a distribution (such as the median or the first Trimmed L-moment of order n), a more realistic and safer approach in a portfolio choice framework would probably be to stick with the first conventional moment, since it is more in line with the value of the portfolio, as directly perceived by the investor and recommended by market authorities. The third and last reason is that replacing mean by median in our preliminary tests, did not lead to huge differences in our application; that is to say that the main general conclusions of the following empirical study (see Section 4) stay the same in our long-only plain vanilla stock application (see also Footnote 16).

$$\mathfrak{M} = \left\{ \lambda_{\mathbf{w}} \in \mathfrak{D} : \forall\, \tilde{\lambda} \in IR^4,\ [(\tilde{\lambda} \succeq \lambda_{\mathbf{w}}) \text{ and } (\tilde{\lambda} \neq \lambda_{\mathbf{w}})] \Rightarrow \tilde{\lambda} \notin \mathfrak{D} \right\}. \tag{2.31}$$

The strong efficient portfolio frontier is then the set of portfolios, defined by their weights $\mathbf{w}$, such that the associated L-moment quadruplet is not strictly dominated in the four-dimensional space. It is then given in the simplex by:

$$\mathfrak{E} = \{ \mathbf{w} \in \mathfrak{A} : \lambda_{\mathbf{w}} \in \mathfrak{M} \}. \tag{2.32}$$

By analogy with tools developed in the field of the production theory (see Luenberger (1995)), the next section introduces the so-called shortage function as an indicator of a portfolio L-moment (in)efficiency, and presents the non-convex higher-order L-moment version of the portfolio optimization program. The solution of the resulting program will be called the four L-moment Efficient Set.

2.3.4 The Shortage Function and the Robust Efficient Frontier

In order to obtain the set of portfolios of the weakly efficient frontier, we need to resolve a multi-objective optimization problem. That is maximize simultaneously the first and the third order L-moments and minimize the second and the fourth L-moments. Several methods allowing the solution of multi-objective problems have been proposed in the literature. Goal Programming, a branch of multi-objective optimization theory introduced by Charnes, Cooper and Ferguson (1955), operates with a set of linear objective functions. Since higher L-moments are clearly non-linear, such an approach should be banned in our case. Another intuitive approach is to aggregate all objectives in a global weighted target function. Optimal portfolios could then be obtained using a traditional non-linear optimizer, but one still needs to specify the importance of the different objectives, which is finally equivalent to introducing a utility function into the problem.

In the following, we choose to use a sequential quadratic programming method to solve our problem. The introduction of a shortage function enables us to optimize si-

multaneously all the objectives, since this latter function measures the distance between some points of the possibility set and the efficient frontier (see Luenberger (1995)). The properties of the set of the portfolio return moments on which the shortage function is defined have already been discussed in the mean-variance plane by Briec, Kerstens and Lesourd (2004) and in the higher moment space by Jurczenko, Maillet and Merlin (2006), Ryoo (2007), Briec and Kerstens (2007), Briec, Kerstens and Jokung (2007), and Yu, Wang and Lai (2008). Their definitions can be extended to obtain a portfolio efficiency indicator in the four L-moment framework. The shortage function associated to any portfolio $\mathbf{w}$ in the feasible set $\mathfrak{A}$, with reference to the direction vector $\mathbf{g} = (\, g_1 \,;\, g_2 \,;\, g_3;\, g_4)$, with $\mathbf{g} \in (IR_+ \times IR_- \times IR_+ \times IR_-) \setminus \{0\}$, in the mean-L-variance-L-skewness-L-*kurtosis* space, is the real-valued function $S_{\mathbf{g}}(.)$ defined such as:

$$S_{\mathbf{g}}(\mathbf{w}) = \underset{\delta \in IR_+}{\text{Sup}} \{\delta : (\lambda_{\mathbf{w}} + \delta \mathbf{g}) \in \mathfrak{F}\}. \tag{2.33}$$

We have the following existence result regarding the shortage function.

Proposition 1. *For every* $\mathbf{g} \in (IR_+ \times IR_- \times IR_+ \times IR_-) \setminus \{0\}$ *and every* $\mathbf{w} \in \mathfrak{A}$*, there exists a unique element* δ^**,* $\delta^* \in IR_+$*, such that:*

$$S_{\mathbf{g}}(\mathbf{w}) = \lambda_{\mathbf{w}} + \delta^* \mathbf{g}. \tag{2.34}$$

See Appendix 2 for proof.

The use of the shortage function in the mean-L-variance-L-skewness-L-*kurtosis* can unfortunately only guarantee the weak efficiency for a portfolio since it does not exclude projections on the vertical and horizontal parts of the frontier allowing portfolios for additional improvements, but hopefully constraints can be easily imposed in the practical implementation for searching only strong efficient portfolios[22] (see below).

The disposal representation of the feasible portfolio set can now be used for deriving

[22]Wierzbicki (1986) proposes a theorem of characterization of strong efficient solutions (based on the evaluation of marginal substitution rates between objectives) that allows us to remove a part of weak efficient portfolios from the set of optimal solutions. However, we did not want at this stage to impose any explicit preference specification in a particular utility function setting.

the lower bound of the true but unknown four L-moment efficient frontier, through the computation of the associated portfolio shortage function. Let us consider a specific portfolio p, defined by its vector of weights denoted $\mathbf{w}_p$, compound from a set of N assets and whose performance needs to be evaluated in the four L-moment dimensions. We then define the function $\Phi_{\mathbf{w}_p,\,\mathbf{g}}(.)$ from $\mathfrak{A}$ to IR_+ by:

$$\Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}) = \underset{\delta \in IR_+}{\text{Sup}} \left\{ \delta : \lambda_{\mathbf{w}} \succeq \left(\lambda_{\mathbf{w}_p} + \delta \mathbf{g} \right) \right\}. \tag{2.35}$$

We also remark that:

$$\Phi_{\mathbf{w}_p,\mathbf{g}}(\mathbf{w}) = \underset{i \in [1,\ldots,4]}{\text{Min}} \left\{ \left[\lambda_i \left(R_{\mathbf{w}} \right) - \lambda_i \left(R_{\mathbf{w}_p} \right) \right] (g_i)^{-1} \right\}, \tag{2.36}$$

where g_i is the *i-th* component, with $i = [1, \ldots, 4]$, of the direction vector $\mathbf{g}$.

The function $S_{\mathbf{g}}$ is related to the function $\Phi_{\mathbf{w}_p,\,\mathbf{g}}(.)$ by the following relation:

$$S_{\mathbf{g}}(\mathbf{w}_p) = \underset{\mathbf{w} \in \mathfrak{A}}{\text{Sup}} \left\{ \Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}) \right\}. \tag{2.37}$$

Using the Goal Attainment method (see Gembicki and Haimes (1975) for a general comprehensive presentation), the shortage function for this portfolio is then computed by solving the following non-linear optimization program $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}}$:

$$\mathbf{w}^* = \underset{(\mathbf{w},\,\delta) \in (\mathfrak{A} \times IR_+)}{\text{ArgMax}} \left\{ \Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}) \right\}, \tag{2.38}$$

where $\mathbf{w}^*$ is a $(N \times 1)$ weakly efficient portfolio weight vector that (weakly) maximizes the expected performance, L-variance, L-skewness, and L-*kurtosis* relative improvements over the evaluated portfolio p in the direction vector $\mathbf{g}$.

Using the vectorial notations of the portfolio return higher L-moments in (3.23) and using the first four L-moments of the specific evaluated portfolio $\mathbf{w}_p$ in the expression of the direction vector $\mathbf{g}$, the non-parametric portfolio optimization program (2.38) can then

be written in a restated version[23] such as (with previous notations):

$$\mathbf{w}^* = \underset{(\mathbf{w},\,\delta)\in(\mathfrak{A}\times\, IR_+)}{\text{ArgMax}} \{\delta\}$$
$$s.t. \begin{cases} \lambda_1\left(R_{\mathbf{w}_p}\right) + \delta\,\lambda_1\left(R_{\mathbf{w}_p}\right) \leq \mathbf{w}'\mathbf{E} \\ \lambda_2\left(R_{\mathbf{w}_p}\right) - \delta\,\lambda_2\left(R_{\mathbf{w}_p}\right) \geq \mathbf{w}'\,\mathbf{\Omega}_{\mathbf{w}}^{(L)} \\ \lambda_3\left(R_{\mathbf{w}_p}\right) + \delta\,\lambda_3\left(R_{\mathbf{w}_p}\right) \leq \mathbf{w}'\,\mathbf{\Sigma}_{\mathbf{w}}^{(L)} \\ \lambda_4\left(R_{\mathbf{w}_p}\right) - \delta\,\lambda_4\left(R_{\mathbf{w}_p}\right) \geq \mathbf{w}'\,\mathbf{\Gamma}_{\mathbf{w}}^{(L)} - 3\,(2)^{-1}\,\mathbf{w}'\,\mathbf{\Omega}_{\mathbf{w}}^{(L)}, \end{cases} \tag{2.39}$$

with:

$$\mathbf{g} = \left[\lambda_1\left(R_{\mathbf{w}_p}\right); -\lambda_2\left(R_{\mathbf{w}_p}\right); \lambda_3\left(R_{\mathbf{w}_p}\right); -\lambda_4\left(R_{\mathbf{w}_p}\right)\right]'.$$

Due to the non-convex nature of the optimization program, we still need to establish the necessary and sufficient conditions showing that a local optimal solution of (2.38) is also a global *optimum*. We actually use the following result.

Proposition 2. *If* $(\mathbf{w}^*,\,\delta^*) \in (\mathfrak{A}\times IR_+)$ *is a local solution of the following non-linear optimization program* $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}}$:

$$\underset{(\mathbf{w},\,\delta)\in(\mathfrak{A}\times\, IR_+)}{\text{Max}} \Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}), \tag{2.40}$$

it is then a global solution.

See Appendix 2 for proof.

Indeed, despite the non-convex nature of the first four L-moment portfolio selection program, the shortage function maximization achieves a global optimum for the non-linear portfolio optimization program. This makes the shortage function technique superior to the other primal and dual approaches of the four moment efficient set, since the latter only

[23]However, restricting (for instance) the search in the direction of an increase in the expected return may lead us to miss some peculiar portfolios that exhibit low (negative) expected returns but have other advantageous characteristics (namely low volatility, outstanding high skewness and rather small *kurtosis*). Nevertheless, in the specific context of portfolio selection, it is doubtful that such portfolios (lotteries) might be considered by any rational investor as optimal. We thus restrict our study to positive expected returns in the algorithm implementation that follows. The same problem arises when looking for a systematic increase in the skewness. In this case, we have to authorize the third coordinate of vector $\mathbf{g}$ to be zero.

guarantees to end with a local *optimum*. In the next section, we illustrate the shortage function technique in the case of a robust strong efficient portfolio selection.

For obtaining the set $\mathfrak{M}$, which corresponds to the set of portfolios whose first four L-moments are not simultaneously dominated, we then consider the evolutionary optimization problem $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}^j}$ at step j, $j \in IN^*$, such as (with previous notations):

$$\begin{aligned} \mathbf{w}^* \;=\; & \underset{(\mathbf{w},\,\delta)\in(\mathfrak{A}\times\, IR_+)}{\text{ArgMax}} \{\delta\} \\ & s.t. \begin{cases} \Delta_{1,\,\mathbf{w}_p}(\mathbf{w},\delta,\mathbf{g}^j) \leq 0 \\ \Delta_{2,\,\mathbf{w}_p}(\mathbf{w},\delta,\mathbf{g}^j) \geq 0 \\ \Delta_{3,\,\mathbf{w}_p}(\mathbf{w},\delta,\mathbf{g}^j) \leq 0 \\ \Delta_{4,\,\mathbf{w}_p}(\mathbf{w},\delta,\mathbf{g}^j) \geq 0, \end{cases} \end{aligned} \tag{2.41}$$

where $\Delta_{i,\,\mathbf{w}_p}(.)$, with $i = [1, ..., 4]$, being the following admissible portfolio directional differences:

$$\begin{cases} \Delta_{1,\,\mathbf{w}_p}(\mathbf{w},\,\delta,\,\mathbf{g}^j) = \lambda_1\left(R_{\mathbf{w}_p}\right) + \delta g_1^j - \mathbf{w}'\mathbf{E} \\ \Delta_{2,\,\mathbf{w}_p}(\mathbf{w},\,\delta,\,\mathbf{g}^j) = \lambda_2\left(R_{\mathbf{w}_p}\right) + \delta g_2^j - \mathbf{w}'\,\mathbf{\Omega}_{\mathbf{w}}^{(L)} \\ \Delta_{3,\,\mathbf{w}_p}(\mathbf{w},\,\delta,\,\mathbf{g}^j) = \lambda_3\left(R_{\mathbf{w}_p}\right) + \delta g_3^j - \mathbf{w}'\,\mathbf{\Sigma}_{\mathbf{w}}^{(L)} \\ \Delta_{4,\,\mathbf{w}_p}(\mathbf{w},\,\delta,\,\mathbf{g}^j) = \lambda_4\left(R_{\mathbf{w}_p}\right) + \delta g_4^j - \mathbf{w}'\,\mathbf{\Gamma}_{\mathbf{w}}^{(L)} + 3\,(2)^{-1}\,\mathbf{w}'\,\mathbf{\Omega}_{\mathbf{w}}^{(L)}, \end{cases}$$

with g_i^j is the *i-th* component of the direction vector $\mathbf{g}$ at step j.

For illustration purposes, we start to solve the problem considering a portfolio p with a first simple direction function $\mathbf{g}^1 = (g_1^1,\, g_2^1,\, g_3^1,\, g_4^1)$, such as $\mathbf{g}^1 = (1, -1, 1, -1)$. Let $(\mathbf{w}^1,\, \delta^1)$ be a solution of $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}^1}$ and let $S^1 = \left\{i = [1, \ldots, 4] :\; \Delta_{i,\,\mathbf{w}_p}(\mathbf{w}^1,\, \delta^1,\, \mathbf{g}^1) = 0\right\}$ be the set of indexes of saturated constraints. If not all constraints are saturated, *i.e.* $S^1 \neq \{1, \ldots, 4\}$, then we consider the optimization problem $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}^2}$, where $\mathbf{g}^2$ is defined by $g_i^2 = g_i^1$ if $i \notin S^1$, and $g_i^2 = 0$ if not. For a solution $(\mathbf{w}^2,\, \delta^2)$ of the problem $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}^2}$, if all constraints are saturated, then the portfolio defined by $\mathbf{w}^2$ is strongly efficient. Otherwise, we consider the new problem $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}^3}$ and we continue in the same manner until all constraints are saturated.

The idea of this optimization process is to saturate at each step at least one of the four constraints, whilst keeping saturated the already saturated constraints. In this aim, at each step, the starting point is the one obtained at the previous step, and the direction function is modified in order for the optimization process to follow a path along the weak efficient frontier. When all constraints are saturated, we then obtain a strongly efficient solution $\mathbf{w}^*$, *i.e.* a set of moment $\lambda_{\mathbf{w}*}$ that belongs to efficient portfolio set $\mathfrak{M}$.

We have here chosen for the starting direction function $\mathbf{g}^1 = (1, -1, 1, -1)$ for the sake of simplicity. Indeed, with such a fixed vector, the whole strong efficient frontier will be obtained by considering different starting points. However, for completing (and boosting) the optimization process, different starting points (portfolios p) and different first direction functions $\mathbf{g}^1$ in $(0; a] \times [-b; 0) \times (0; c] \times [-d; 0)$ are considered in the following practical implementation. The values of a, b, c and d are set in realistic ranges of potential improvements[24] and take into account the differences of scale between the four L-moments.

2.4 Data and Empirical Results

In the following empirical application, we explore a dataset of quotes of some of the most liquid European stocks, provided by *Bloomberg*, in the period from June 2001 to June 2006. The database consists of weekly Euro denominated returns of a sample of 162 stocks included in the *DJ European Stoxx* index. First, the selected stocks were not chosen randomly as in some previous studies, but were selected for obtaining a cylindric sample. The 162 stocks considered (representing approximately more than 1.3 trillion of Euros in terms of free float market capitalization at the end of the sample[25]) consist of all stocks present over the whole sample period that have not experienced a corporate action (such as a stock split for instance) during the sample period. Secondly, as in Jondeau and Rockinger (2006), we have chosen a weekly sample frequency. Indeed, it is worth emphasizing here that the frequency of data is often claimed to affect both departures

[24]These parameters are fixed hereafter to the maximum values of the L-moments found on individual assets in the sample.

[25]that corresponds to a quarter or so of the total capitalization of the European index.

from normality and serial correlation patterns of returns and volatilities. In our case, the frequency of observations[26] has been chosen in order to be low enough for being well adapted to an asset allocation problem (in which reallocations cannot happen very often), but high enough for keeping in the sample the main peculiarities of the financial returns, such as the skewed-heteroskedasticity *phenomenon*, that generally goes with unconditional asymmetric and leptokurtic underlying return distributions. Thirdly, the period of study is also rich in events (end of the internet bubble crash, the 9/11/2001 event, the market correction of May 2006...) and is characterized by a bear market on the first part of the sample (2001-2003), followed by a strong bull market (2003-2006). If we note that market performance is very high on the total period (with an annualized return of 18% for the *DJ European Stoxx* index, whilst the typical annualized return on the American stock DJI is equal to 5% on the period 1900-2006), we also remark that we have various market conditions (rallies, bear markets, booms and crashes) in the sample period (which is similar to those studied on the American stock market by Maringer and Parpas (2007)); it allows us to think that the sample is not too specific for our general purpose.

Since our aim is to evaluate whether, in some instances, the widely-used mean-variance *criterion* may be inappropriate in selecting the optimal portfolio weights, we shall check, before all, the univariate non-Gaussianity of the sample stock return series, using main classical Normality tests, namely Jarque-Bera, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. The Jarque-Bera test is one of the most used portmanteau Goodness-of-Fit measures of departure from normality and is based on sample skewness and *kurtosis*. The statistic of the test has an asymptotic Chi-squared density; however, it has been proven to have limited power in a small sample, because empirical counterparts of conventional third and fourth moments approach Gaussianity only very slowly. The second test we performed is the Kolmogorov-Smirnov one, which is another main classical Gaussianity test; it is based on the observed largest difference between the data-driven cumulative Empirical Density Function and the sample estimate of the Normal reference distribution.

[26] In Jurczenko, Maillet and Merlin (2006), a monthly frequency (on hedge funds) was used, and first preliminary tests made here with daily returns (on stocks) showed no difference in overall general results of this article.

Correcting for the bias of using sample estimates of characteristic parameters of the reference law leads to the third test considered, which was proposed by Lilliefors. This test is designed with the null *hypothesis* that data come from a normally distributed population, when the tested *hypothesis* does not specify which normal distribution (*i.e.* without specifying expected value and variance). One of the peculiarities of this test is to be not too sensitive to outliers and thus more sensible to the adequation of the central part of the distribution. Finally, we also used the Anderson-Darling test, which is known for being one of the most powerful statistics for detecting departures from normality (see Stephen (1974), d'Agostino and Stephens (1986)). This test is based upon the concept that when given a hypothesized underlying distribution, the data can be transformed into a Uniform distribution, and is then crucially linked to tails of the data density.

We thus began by testing the effective (non-)normality of our stock return sample, focusing on different aspects of Gaussianity: using an explicit test on both skewness and *kurtosis* (according to Jarque-Bera tests), testing the largest inadequations (based on Kolmogorov-Smirnov tests), emphasizing the differences in centers of distributions (with Liliefors tests), as well as discrepancies in the tails (defined through Anderson-Darling tests). Not surprisingly, the vast majority of the original stock return series cannot indeed be considered as Gaussian (see below).

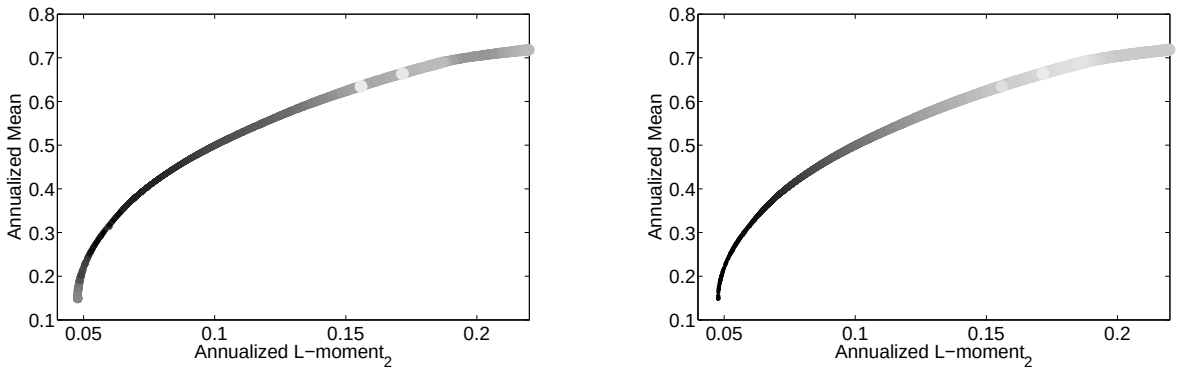
We then start the efficient portfolio search algorithm by generating randomly one thousand arbitrary portfolios, and then optimize each one in their four moments using our distance function approach in 9,637 directions[27]. We present hereafter the empirical four-moment efficient frontiers and their projections in the various planes. Due to the large number of stocks considered in the efficient portfolios, this optimization problem belongs to a large-scale asset allocation problem class (see Perold (1984)). The main consequence is that we observe some strong discontinuities in the empirical efficient frontiers. This last feature is also intensified by the addition of strong short-sale constraints[28].

[27]It corresponds to the number of combinations of 11 possible intensities (from 0 to 1 with a .1 step for each L-moment).

[28] These constraints are realistic in the case of pure equity portfolios, but can easily be relaxed, to some extent, to relative constraints to the exposed capital as in Pástor and Stambaugh (2000) and So and Tse (2001).

In the following, we will first start by evaluating the trade-offs between each pair of moments; we secondly analyze the efficient portfolios and their global characteristics, paying special attention to the higher moments that are neglected in the traditional analysis. In all the following representations, we further restrict the efficient portfolio set in considering only those with a positive mean (in a portfolio selection context) and a reasonable second annualized volatility[29]. After having presented some efficient portfolio frontiers, we thirdly evaluate the optimality of (primal) potential efficient portfolios when grouping portfolios together based on their similarities in terms of L-moments, then when valuing them according to a specific utility function in the dual approach. Fourthly, we will make the higher moments of efficient portfolios more appealing, transforming the real third and fourth L-moments to intensify potential distortions and evaluate the impact on optimal asset allocations. Finally, we will see that almost similar portfolios in terms of first and second L-moments may exhibit some differences in higher moments that might be, for the investor, interesting to control.

Figure 2.2. First Four L-moment Constrained Efficient Frontier in the L1-L2 Plane.



with Third L-moments Coloured

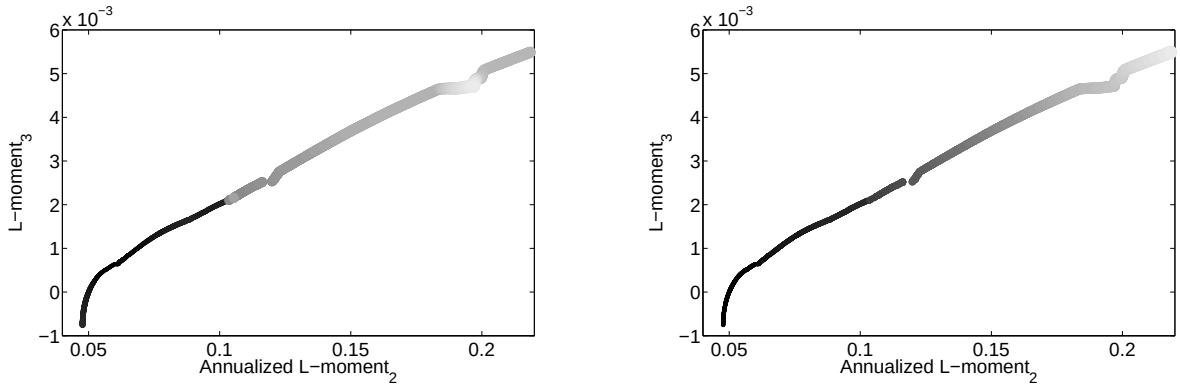
with Fourth L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 2,610 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the third (fourth) L-moments on the left figure (the right figure).

[29]The maximum annualized volatility level considered here is 40%, corresponding to a L-moment$_2$ inferior to .22. This restriction allows us to avoid representing the distortions caused by the no-short sale constraints for portfolios with high levels of volatility.

The Figures 2.2 to 2.7 represent the frontiers of the projections of the four L-moment efficient portfolios in two specific L-moment dimensions. In Figure 2, the traditional two-moment frontier (see Markowitz (1952)) is represented in its robust version. We find exactly the same type of shape for this restricted frontier: the relation between the mean and the L-moment$_2$ is a non-linear increasing one. If investors want to target better mean returns, they have to accept a higher dispersion of returns. When we look at the other characteristics of these efficient portfolios, we see that some high skewness portfolios appear for mild medium means and reasonable second L-moments (see Figures 2.2 and 2.3).

Figure 2.3. First Four L-moment Constrained Efficient Frontier in the L2-L3 Plane.



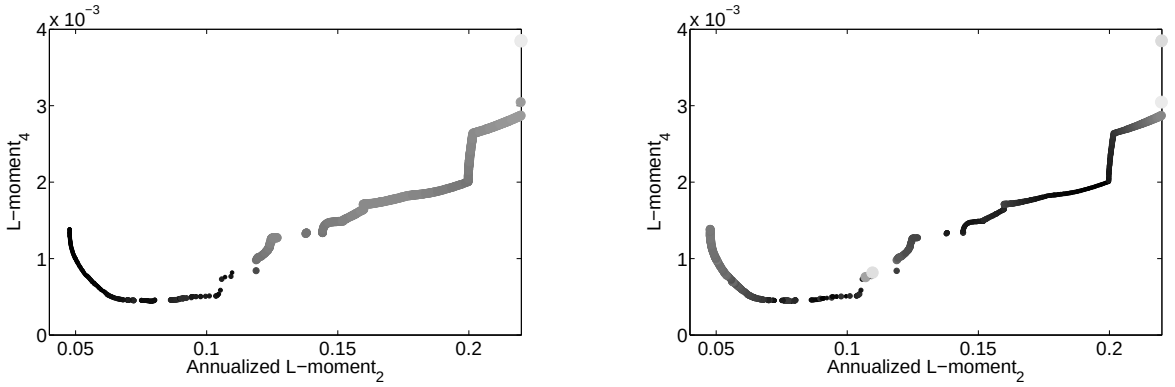
with First L-moments Coloured

with Fourth L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 1,932 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the first (fourth) L-moments on the left figure (the right figure).

We also remark that the second and fourth L-moments seem to vary in the same direction. This is confirmed in Figure 2.4, in which we represent the lowest fourth L-moments of the four L-moment efficient portfolios for each level of the L-moment$_2$. Thus, the value of the L-moment$_2$ of efficient portfolios depends positively on the level of the first and fourth L-moments. Accordingly, when investors tend to reach higher expected returns, they have to face an increase of the L-moment$_2$ and the L-moment$_4$ (see also Figure 2.5). In other words, increases of L-moment$_4$ go with increases of L-moment$_2$ for efficient portfolios. This result

Figure 2.4. First Four L-moment Constrained Efficient Frontier in the L2-L4 Plane.



with First L-moments Coloured

with Third L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 731 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the first (third) L-moments on the left figure (the right figure).

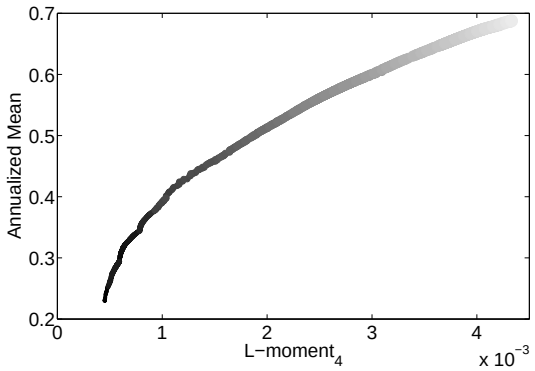
is consistent with those by Maringer and Parpas (2008)[30], and seems natural and intuitive when considering the traditional interpretation of L-moment$_2$ in terms of dispersion of returns and of the L-moment$_4$ as a consequence of the presence of extreme returns (that, *ceteris paribus*, entails a higher dispersion measure). The two positive relations (between the first and the second, the second and the fourth L-moments of four-moment efficient portfolios) impose a clear positive link between the first and fourth L-moments, as seen in Figure 2.6: the investors are here compensated for bearing more extreme risks.

Regarding now the relations between L-moment$_3$ and others L-moments for efficient portfolios, it seems far less clear as already seen in Figure 2.2 and indicated in the literature on the impact of skewness on the expected utility (see Brockett and Kahane (1992), Brockett and Garven (1998), and Christodoulakis and Peel (2006)). Moreover, frontiers present large discontinuities as already pointed out by Athayde and Flôres (2002)[31]. In Figure 2.6, the highest mean efficient portfolios for each level of L-moment$_3$ are presented.

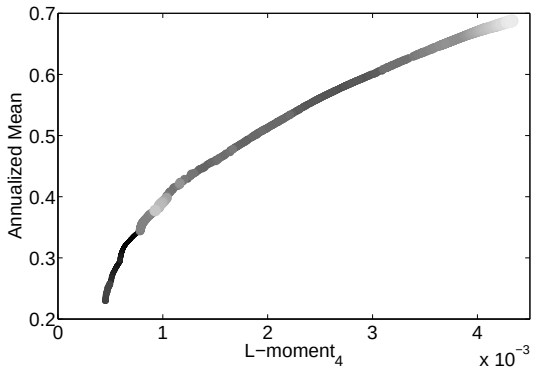
[30]despite exhibiting an apparent inverse relation since they used the fourth standardized moment. By construction, standardizing the fourth moment necessarily leads to an inverse relation with volatility.

[31]The discontinuities are here reinforced by the imposed no-short sale constraint, and also, to an even greater extent, by the fact that we perform here a large scale optimization (see Athayde and Flôres (2002)).

Figure 2.5. First Four L-moment Constrained Efficient Frontier in the L1-L4 Plane.



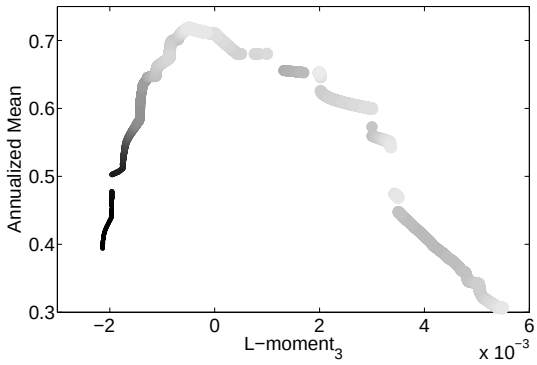
with Second L-moments Coloured



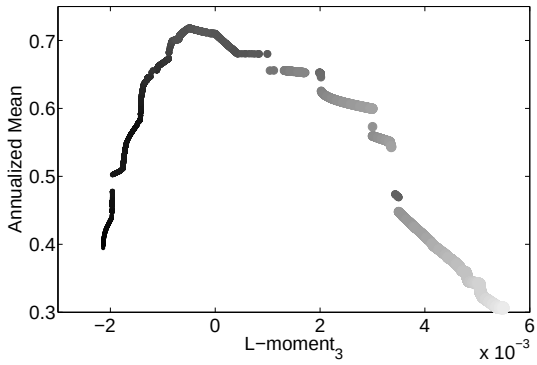
with Third L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 731 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the second (third) L-moments on the left figure (the right figure).

Figure 2.6. First Four L-moment Constrained Efficient Frontier in the L1-L3 Plane.



with Second L-moments Coloured



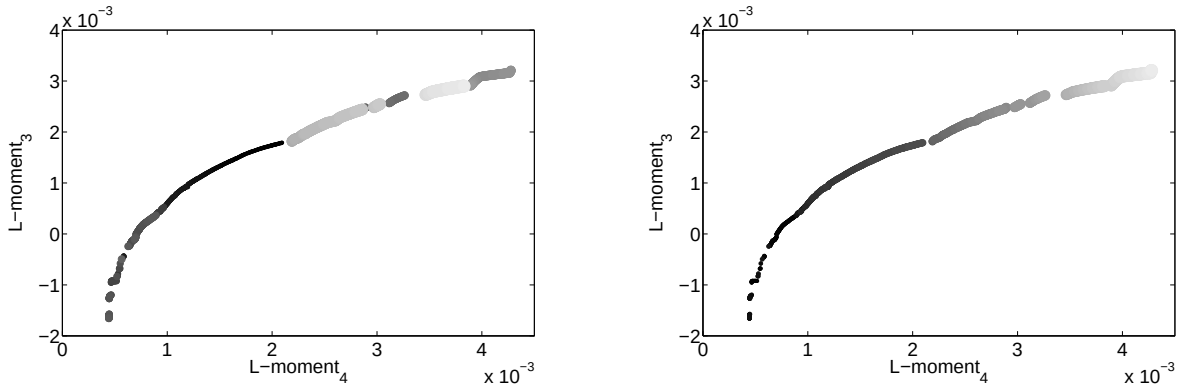
with Fourth L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 1,925 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the second (fourth) L-moments on the left figure (the right figure).

We first note that the maximum mean return portfolio has a negative skewness. That is: willing high mean returns has a price in terms of L-moment$_3$ (called a skewness *premium*; see Post, Vliet and Levy (2008) and relevant literature). From Figure 2.6 (and Figure

2.2), it also appears that the highest L-moment$_3$ efficient portfolios have medium mean returns (and L-moment$_2$). Finally, the lowest mean (L-moment$_2$) efficient portfolios are associated with a low (negative) asymmetry as shown in Figure 3 and Figure 6, the former representing the link between the highest L-moment$_3$ for various L-moment$_2$ of the four L-moment efficient portfolios. An interesting feature for the investor is revealed in this figure: for some quite low volatility portfolios, the third L-moment could be strongly improved with second (and fourth) L-moment(s) only marginally deteriorated. The non-linear non-monotonic discontinuous relations found between L-moment$_3$ and other L-moments for efficient portfolios (see also Figure 2.7) are consistent with the relations presented by Athayde and Flôres (2004), Jurczenko, Maillet and Merlin (2006), and Maringer and Parpas (2008), but differ from the relation found in Harvey, Liechty, Liechty and Müller (2002)[32].

Figure 2.7. First Four L-moment Constrained Efficient Frontier in the L3-L4 Plane.



with First L-moments Coloured with Second L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 1,244 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the first (second) L-moments on the left figure (the right figure).

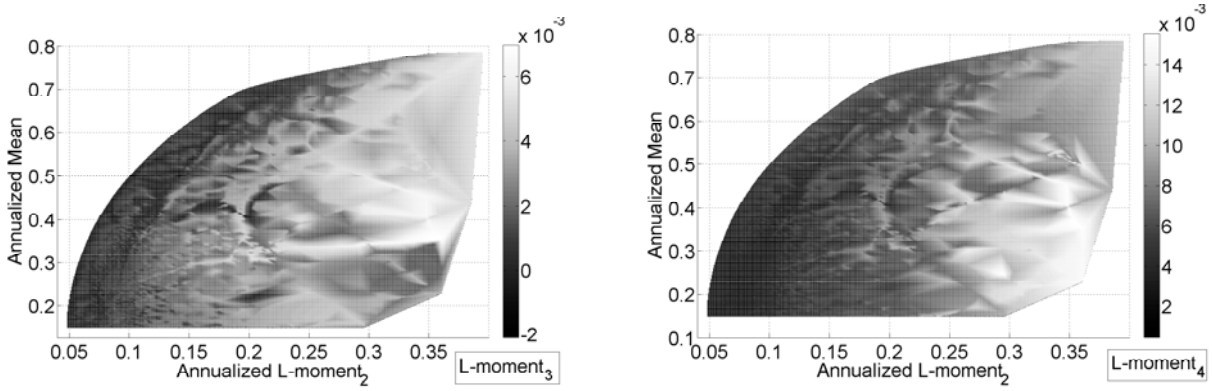
The investigation of the various two-moment trade-offs between L-moments leads to the following preliminary conclusions. First, the strong link found between the even L-

[32]This difference could be explained by a database bias (four equity stocks are used and one of them dominates the others in the first and third moments).

moments seems to be in accordance with the traditional two-moment analysis and can be thought of as rational (even if not automatic - see Brockett and Garven (1998), and Haas (2007)). Since the even L-moments of efficient portfolios are thus found to be in correlation, the temperance of agents (sensitivity to the fourth moment) is closely linked with their risk aversion (sensitivity to the second moment). As seen in Jurczenko and Maillet (2004) in the context of option pricing with higher-moments and as mentioned by Galagedera and Maharaj (2008) examining an extended higher comoments conditional CAPM with realized variables, the (co-)*kurtosis* seems not to bring a lot of information into the analysis when the (beta) second moment is already in (at least for the type of assets and data sample considered here). Secondly, as already pointed out in the literature, the relation between odd moments is proved unclear. Indeed, the relation between first and third moments for the efficient portfolios is shown to be non-monotonic (see Post, Vliet and Levy (2008), and Crainich and Eeckhoudt (2008)). Extreme (highest or lowest) greediness of investors (highly sensitive to the first moment) goes with a lower prudence (sensitivity to the third moment), whilst extremely prudent agents will choose portfolios with mild (non extreme) characteristics in terms of first, second and fourth moments. Thirdly, as a consequence, we may say that the prudence characteristic of the investor is crucial when determining her asset allocation, and that taking into account the third-moment might lead to favorable improvements of the utility of the investor, only causing small changes in the three other moment characteristics of optimal portfolios. We further investigate, hereafter, these preliminary conclusions, not only dealing with moments two-by-two as previously, but taking into account all moments at the same time, and focusing more on higher moments than in the above analysis.

We represent in Figure 2.8 the whole set of the four-moment efficient portfolios in the traditional plane of Markowitz. All portfolios in this Figure are said to be (potentially) Pareto efficient since one may find a rational agent, exhibiting a specific set of marginal rates of substitution between L-moments, that chooses one of these portfolios without any hope of simultaneous improvements in all moments going with another choice. All these portfolios, which do not belong to the Mean-L-moment$_2$ frontier anymore, show either

Figure 2.8. First Four L-moment Constrained Efficient Frontier in the L1-L2-L3 Space.



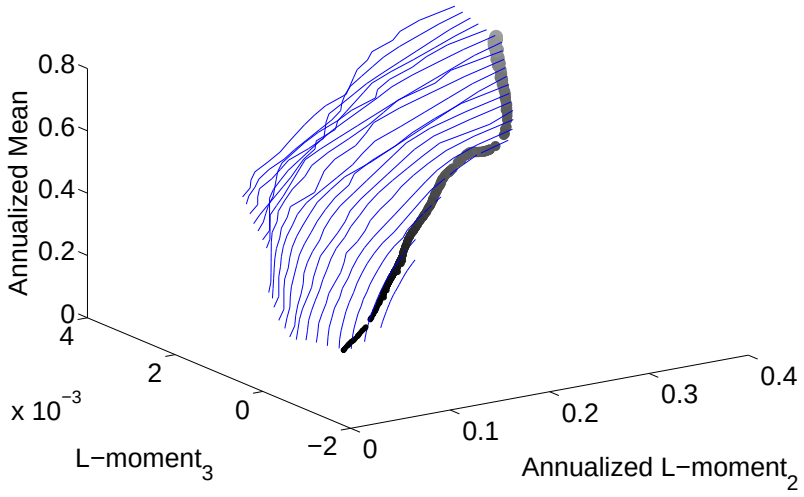
with Third L-moments Coloured with Fourth L-moments Coloured

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 65,253 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent (respectively) the third (fourth) L-moments on the left figure (the right figure).

a high L-moment$_3$ or a low L-moment$_4$ amongst all possible portfolios, or a compromise between the two. Color shadings represent either the third or the fourth L-moments respectively on the left and right sub-figures. Three pieces of complementary information to the previous analysis should be mentioned here. First, almost all the interior of the two-moment frontier is full of four-moment efficient portfolios, except the right bottom part of the Figure 2.8 that does not contain any optimal portfolios: almost any compromise between moments is here reachable, except low mean returns accompanied by high second, third and fourth moments (which are dominated by a better mean portfolio). Secondly, we do not observe a clear link between moments when enlarging the analysis in the four dimensions. For instance, minimization of the second L-moment from high levels (right part of the figure) results in various levels of other moments. Thirdly, and surprisingly, agents who are extremely prudent (paying high attention to the third L-moment) will choose portfolios in the same region (characterized by high-mean, high-volatility, high-skewness and also high-*kurtosis* portfolios) than the agents having the lowest temperance (almost insensitive to the *kurtosis*). But we also remark that choosing amongst these portfolios represents a large (probably unrealistic) price in terms of expected return. In

that sense, diverging too far from the two-moment only optimization paradigm (looking for interesting higher moments) has a cost, which may be seen as almost unreasonable by most investors. However, and once again, we also see portfolios, near the traditional frontier (upper limb), that exhibit better third L-moments. The third and fourth moments of the Mean-L-moment$_2$ frontier portfolios could thus probably be improved, with only a marginal deterioration of the first two L-moments.

Figure 2.9. First Four L-moment Constrained Efficient Frontier in the L1-L2-L3 Space.



Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The constrained efficient frontier is compounded with 65,253 simulated portfolios, obtained after optimization in 9,637 directions of random portfolios built with 162 European equities. Lines correspond to Mean L-moment2 Efficient Portfolios when the level of L-moment3 is constrained. Sizes of dots and their colour shadings (from black small low levels to white large high levels) represent the Four L-moments of Efficient Portfolios in the Mean-L-moment2 plane (i.e. the traditional Efficient Portfolios).

We highlight this last fact through Figure 2.9 where we plot the frontiers of Mean-L-moment$_2$ efficient portfolios for various given levels of L-moment$_3$, together with the global mean-L-moment$_2$ efficient frontier. The sizes of the dots (altogether with their color shadings here) represent the level of fourth L-moment. We thus obtain various peculiar shapes of frontiers for each level of L-moment$_3$, that are each time very similar to the two-dimension theoretical *parabola*, with different asymptotes and apexes depending upon the third moment. This result is then comparable to the representations of the Mean-Variance-Skewness presented in Jurczenko and Maillet (2001), Buckley, Saunders

and Seco (2008), and Mencía and Sentana (2008) in a restricted three-moment world, and by Jurczenko, Maillet and Merlin (2006), Kerstens, Mounir and Woestyne (2007), or Maringer and Parpas (2008) in a four parameter space. The L-moment$_2$ efficient portfolios with low volatilities indeed appear to be located on the most negative L-moment$_3$ parts of the shapes. This highlights the fact that for low and high L-moment$_2$ targeted portfolios, L-moment$_3$ and L-moment$_4$ could be significantly improved by reasonable changes in the L-moment$_2$, as shown in Table 3.13 in the case of low volatility portfolios.

Table 2.1. P-statistics of Normality Goodness-of-Fit Tests of the various Optimal Portfolio Return Series (frequencies of rejections at 1%, 5% and 10% confidence levels)

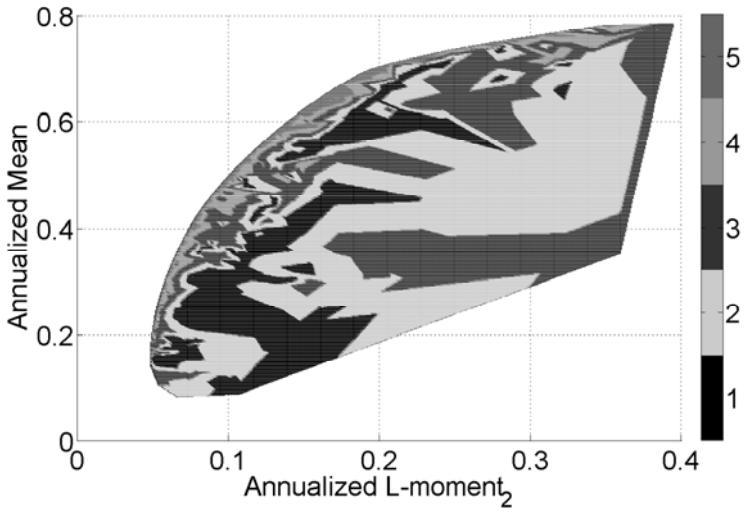
	Jarque-Bera	Kolmogorov-Smirnov	Liliefors	Anderson-Darling
Original European Stock Returns Series	(.91)/[.96]/.98	(.20)/[.38]/.58	(.68)/[.81]/.86	(.82)/[.90]/.94
Four L-moment General Efficient Portfolios	(.65)/[.72]/.75	(.15)/[.38]/.50	(.64)/[.82]/.89	(.78)/[.91]/.94
Four L-moment Quartic Optimal Portfolios	(.88)/[.93]/.95	(.13)/[.35]/.50	(.69)/[.80]/.84	(.74)/[.81]/.83
Four L-moment CRRA Optimal Portfolios	(.69)/[.78]/.82	(.12)/[.15]/.18	(.20)/[.26]/.45	(.46)/[.48]/.56
Four C-moment Power-expo Optimal Portfolios	(.98)/[.99]/1.00	(.10)/[.23]/.43	(.60)/[.79]/.84	(.78)/[.85]/.86
Four L-moment Power-expo Optimal Portfolios	(.83)/[.95]/1.00	(.14)/[.39]/.44	(.55)/[.60]/.63	(.65)/[.76]/.78

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. From the sets of the (162) original return series, the (65,253) general Four L-moment Efficient Portfolios (see Figure 8), the (379) Four L-moment CRRA Optimal Portfolios, the (123) Four L-moment Quartic optimal portfolio, the (92) Four C-moment Power-expo optimal portfolios, the (341) Four L-moment Power-expo Optimal Portfolios (see Figure 11), we compute the Jarque-Bera, the Kolmogorov-Smirnov, the Liliefors and the Anderson-Darling Goodness-of-Fit tests. The table reports the frequency of P-statistics below the usual probability thresholds (1% between parentheses, 5% between brackets and 10% between accolades), *i.e.* the relative number of (stock) portfolio series being non-normal in the sense of the usual Gaussianity tests.

Now comes the question of utility implications of an allocation also considering higher moments in a dual representation of the investor allocation problem, which is a difficult question mainly because of the complexity of the problem of optimizing utility functions[33].

[33]Several issues indeed arise when optimizing utility functions (see Jondeau and Rockinger (2006) and Garlappi and Skoulakis (2008)). First, excessive sensitivities of utility functions to parameters often plague optimization procedures. Secondly, when performing multi-objective optimization, five theoretical conditions (such as independence of the sub-objectives) have to be respected to ensure the existence of an optimum (see Keeney and Raiffa (1993)). Thus, the multi-objective optimization problem applied

Figure 2.10. Clusters of the Four L-moment Optimal Portfolios according to their L-moments in the L1-L2 Plane.



Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The colours represent the clusters of the Four L-moment Efficient Portfolios according to their L-moment similarity. This figure is obtained using a Self-Organizing Maps' classification algorithm on a string of five cells, applied on vectors of (unscaled) L-moments defining the optimal portfolios (see Kohonen (2000), and Appendix 3).

In a very general context, we can start by using the vectors of L-moments of the potential efficient portfolios for defining regions in which the L-moments of efficient portfolios are almost similar (arbitrarily valuing at the same rate the various L-moments since they are of the same order of magnitude). Figure 2.10 indeed represents the location of portfolios classified into five clusters that have been built using a Self-Organizing Maps algorithm (see Kohonen (2000) and Appendix 3), based on the vectors of L-moments. Each region in Figure 2.10 (color shadings on the map) represents a set of similar portfolios in terms of their L-moments; we remark that we can distinguish clear regions of portfolios with similar L-moment characteristics. Cluster 4, for instance, groups together most of the traditional two-moment optimal portfolios, whilst cluster 2 is compounded with the highest skewness and *kurtosis* portfolios. In other words, in such a general framework (with no formal specification in link with an explicit utility function), individuals may pick

to portfolio choice (see Ehrgott, Klamroth and Schwehm (2004)) has to be rigorously defined. Lastly, as shown by Marschinski, Rossi, Tavoni and Cocco (2007), the parameter uncertainty has strong implications in the determination of optimal solutions.

their optimal portfolios in various rather homogeneous regions (corresponding to several preference and risk prototypes), according, first, to the characteristics of the underlying L-moments (Figure 2.10) and, second, to the way they value them (Figure 2.11 below).

In Figure 2.11, we propose to illustrate the optimal choice, now valuing the investors' sensitivities and moments according to some flexible usual representations of utility functions (see Appendix 1). We select and represent, from the set of the optimal Four L-moment Efficient Portfolios, the ones that maximize the level of utility for various parameters in relevant ranges, for a Quartic utility function computed with L-moments (imposing here the sensitivities to moments to decrease with the order of the moment) and a Quartic (CRRA) utility function (for plausible reported values of the risk aversion coefficient - see Jondeau and Rockinger (2006)[34]). We also generalized our results (bottom figures) using a more flexible Power-exponential utility function. In almost all cases, selected efficient portfolios are lying on (or not far from) the traditional two-moment frontier.[35] We indeed converge to the previous reported results concerning the accuracy of a second-moment approximation, since the four-moment *optima* all lie on the traditional two-moment frontier. As in Simaan (1993) these results suggest, furthermore, that the opportunity cost of the static mean-variance investment strategy is empirically irrelevant in our case. In other words, either the considered class of the utility function is not general enough and/or the departures from Gaussianity and non-linearities present in our database on the period are not large enough to significantly impact the traditional asset allocation.

This fact is strengthened by the results presented in Table 3.13 and Figure 2.12. As already mentioned, most of the original stock return series cannot indeed be considered as Normal (see Table 3.13). The same is also observed regarding the optimized portfolio

[34]As an alternative, we also considered a fourth-order Taylor expansion of the CARA (for Constant Absolute Risk Aversion) exponential utility function, that has also been widely studied in the literature. Yet, it should be noticed, as written in Jondeau and Rockinger (2006), that we found the same basic results with both types of utility functions.

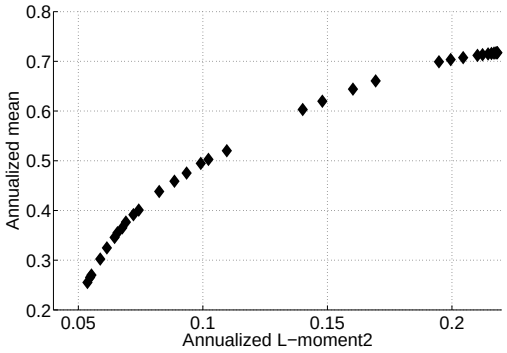
[35] Despite the widespread approach in the investor preference literature being to scale the initial agent wealth to one (see Jondeau and Rockinger (2006)), we also tested *per* curiosity optimal utility-linked choices considering some levels of wealth (for intensifying the role of the higher moments). Our first results were not conclusive in the sense that we found no clear impact on the moment preferences (when sticking to realistic decreasing order sensitivities). Moreover, utility function approximations became in this case far more complicated, since terms appear to concern the dependence upon higher moments (due to the presence of various wealth-moment interactions).

Figure 2.11. Optimal Portfolios for various Four-moment Dependent Utility Functions in the L1-L2 Plane.

CRRA Utility Function with L-moments

$U(W) = (1-\gamma)^{-1} W^{1-\gamma}$

with $\gamma \in]\,0, 50\,[$.

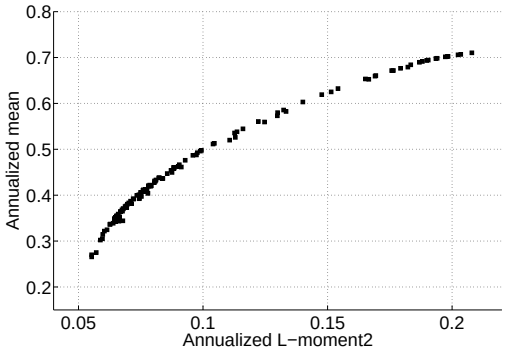


Quartic Utility Function with L-moments

$U(W) = \beta_0 + \beta_1 W + \beta_2 W^2 + \beta_3 W^3 - \beta_4 W^4$

with $\beta_i \in [0, 1] \text{for} i = [1, ..., 4]$

and $\beta_i \geq \beta_{i+1}$ for $i = [1, ..., 4]$.

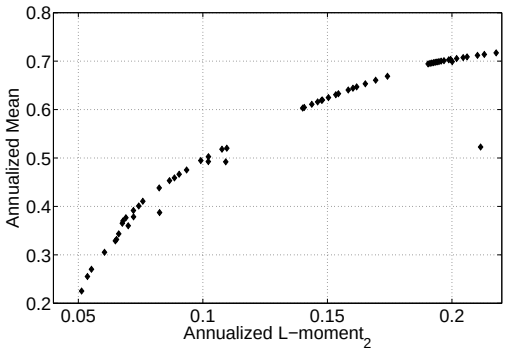


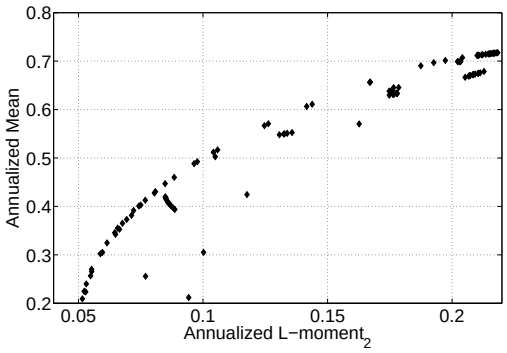
Power-expo Utility Function with Conventional Moments

Power-expo Utility Function with L-moments

$U(W) = \alpha^{-1}\left[1 - \exp\left(-\alpha W^{1-\gamma}\right)\right]$

with $\alpha \in [\,-50, 50\,[\,\setminus\{0\}, \gamma \in [\,-10, 10\,[\,\setminus\{1\}$ and $\alpha(1-\gamma) > 0$.





Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. From the set of directionally optimized portfolios, we select portfolios that are optimal for each of the considered utility functions. In the top left figure, black points represent optimal portfolios when considering CRRA utility functions. In the top right figure, black points represent optimal portfolios associated to Quartic utility functions that respect the condition of decreasing order moment preferences. In the bottom figures, black points represent optimal portfolios with respect to Power-expo utility functions using conventional moments (bottom left figure) and L-moments (bottom right). Consideration of both C-moments and L-moments allows us to ensure that the two frontiers remain in line and to show the impact of the use of robust statistics in portfolio choice.

returns. The Jarque-Bera, Anderson-Darling and Liliefors tests highly reject the Gaussian *hypothesis*, whilst the Kolmogorov-Smirnov test overall result is more contrasted. The two former tests mostly evaluate differences in the tails of the distributions, while the Lilliefors is especially sensitive to distribution gaps located at the node. For the Kolmogorov-Smirnov test, only the size of the largest difference counts. Differences between empirical probability density functions and the theoretical Gaussian-benchmark ones are then mainly differences in the skewness and the *kurtosis*, small differences in the tails and in the centers of distributions, more likely than large differences somewhere specific. It is fair to mention here that most of the series - original stock and various optimized portfolio returns - clearly differ from the ideal Gaussian *hypothesis* of the Markowitz' model. Our negative result concerning the impact of higher moments on the utility-based optimal choice is not due to a hypothetical (almost-)Gaussianity of the underlying series. What could happen now if we were to intensify the original values of the skewness and the *kurtosis* (keeping the two first moments unchanged)? In Figure 2.12, we select and represent, from the set of all directionally optimized portfolios, those that are optimal for one of the various Quartic Utility functions - making their higher moments more attractive by a simple linear transformation (multiplying the third L-moment and dividing the fourth L-moment by 2 and 4, with the first and second L-moments unchanged[36]). In the left (right) figure, the bold grey line represents the Quartic efficient frontier with no change (original data), whilst the thin black line represents the efficient frontier when the third moments of underlying portfolios are multiplied by 2 (respectively by 4), the bold grey dotted line corresponds to the efficient frontier when the fourth moments are divided by 2 (by 4), and the thin black dotted line is related to the efficient frontier when both third and fourth L-moments are (respectively) multiplied and divided by 2 (by 4). We clearly see that, even for unreasonable changes in the higher L-moments (a 100% increase), the optimal portfolios still lie close to the traditional frontier, whilst it is just for a drastic unrealistic intensification that some optimal portfolios start to go away from the usual

[36] Since all the two-moment optimal portfolios exhibit neither the highest L-moment$_3$ nor the smallest L-moment$_4$ for all levels of L-moment$_2$, we also tried other simple linear transformations such as mutliplying by 2 and 4 (dividing by 2 and 4) both higher L-moments, with no significant change in the conclusion.

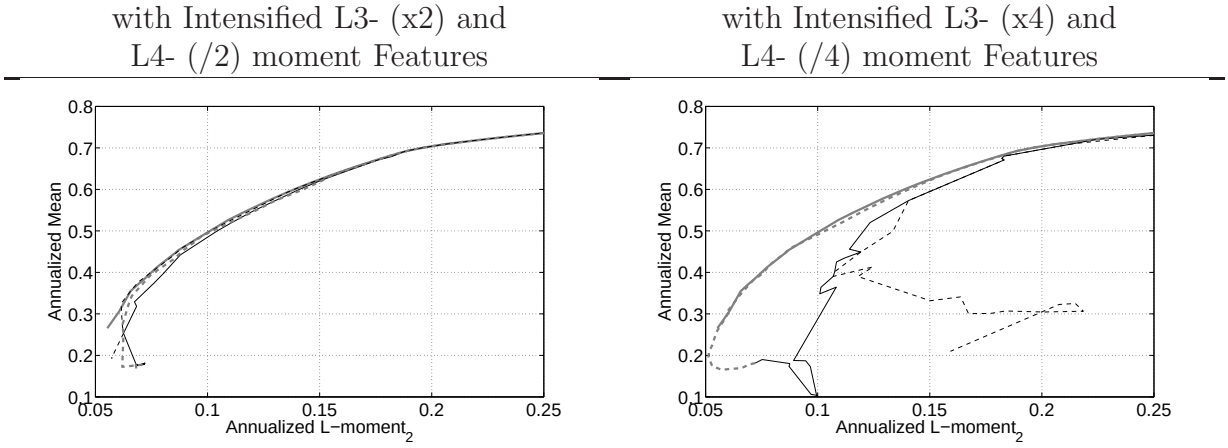
frontier (but only for some low-medium volatility portfolios). This clearly illustrates that even if we greatly intensify the non-normal features of the data, no clear impact could be highlighted in terms of utility improvements.

Figure 2.12. Efficient Frontiers for CRRA Utility Functions with Intensified Higher-order Moments.

CRRA Utility Function with L-moments

$$U(W) = (1-\gamma)^{-1} W^{1-\gamma}$$

with $\gamma \in]\,0, 50\,]$.



Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. From the set of directionally optimized portfolios, we select those that are optimal for one of the various CRRA utility functions. On the left (right) figure, the grey bold line represents the CRRA efficient frontier with no change (original data), whilst the thin black line represents the efficient frontier when the third moments of underlying portfolios are multiplied by 2 (respectively by 4), the bold grey dotted line represents the efficient frontier when the fourth moments are divided[37] by 2 (by 4) and the thin black dotted line represents efficient frontier when third moments are multiplied by 2 (by 4) and fourth moments are divided by 2 (by 4).

This last conclusion, considering our pure equity sample in the period 2001-2006 and some classical higher-moment utility functions, so greatly moderates the previous ones when we only considered the higher moment portfolio characteristics without valuing them through a precise utility function. Whatever indeed the truncation order of the utility function chosen, the accuracy of the approximation of the expected utility is still definitely an empirical issue (see Hlawitschka (1994)). Several authors (such as, for instance, amongst others: Markowitz (1991), and Hsieh and Fung (1999) - see Jurczenko and Maillet (2006a),

and Jondeau and Rockinger (2006), for larger lists of references) show using different assets, databases, time periods, constraints, utility functions and parameter sets, that a second-order Taylor expansion already quite accurately approximates the expected utility, leaving us *a priori* with only small potential improvements when dealing with higher moments (see Jondeau and Rockinger (2003b and 2006)). These studies confirm the previous research on this subject, since we were not able in a traditional expected utility framework to clearly underline the positive effect of incorporating higher-moments in the asset allocation.

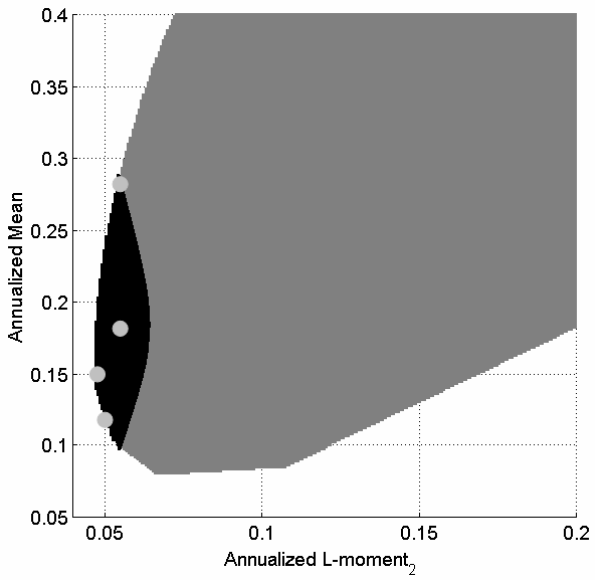
However, a potential application for the investors of the higher L-moments could be in the control of the features of optimal portfolios, since some efficient portfolios, similar in terms of the two first moments, exhibit large differences in higher L-moments as presented in Table 2.2, where we compare some characteristics of second moment alike portfolios. From the set of the Four L-moment Efficient Portfolios with an L-moment$_2$ that is not significantly different from that of the Minimum L-moment$_2$ Portfolio (at a 5% threshold using the Log-normal *hypothesis*[38] by Alizadeh, Brandt and Diebold (2002) - see Figure 2.13), we select the Local Maximum Mean, the Local Maximum L-moment$_3$ and the Local Minimum L-moment$_4$ Portfolios. The Table 2.2 presents the first four L-moments and the Sharpe ratios of the selected local best portfolios and indicates that small (non-significant) changes in the L-moment$_2$ may lead to some differences in other characteristics, advocating for a potential advantageous control of other features of the traditional optimal portfolios.

2.5 Conclusion

We introduce in this article a general method for deriving the set of strong efficient portfolios in the non-convex Four L-moment space, using a shortage optimization function (see Luenberger (1995), Briec, Kerstens and Lesourd (2004), Jurczenko, Maillet and

[38]We first tested the hypothesis of Log-normality of the second L-moment using one-year rolling sample volatility of returns on the global Minimum L-moment$_2$ Portfolio, and the resulting P-statistics of Jarque-Bera, Kolmogorov-Smirnov, Liliefors and Anderson-Darling tests were, respectively, .14, .13, .14, .14, leading to a non-rejection of the null. Using a Fisher test, we then considered the set of portfolios having a return standard deviation not significantly different from the one of the return on the global Minimum L-moment$_2$ Portfolio. We, secondly, considered portfolios non-significantly different in L-moment$_2$, using an empirical cumulative density function; overall results were similar.

Figure 2.13. Set of Portfolios with Second L-moments almost Equivalent to the one of the Minimum L-moment2 Portfolio.



Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The black area represents the region in which the Four L-moment Efficient Portfolios have an L-moment2 not significantly different at a 5% threshold from the L-moment$_2$ of the Minimum L-moment$_2$ (using the Log-normal volatility hypothesis by Alizadeh, Brandt and Diebold (2002) - differences in the Second L-moment are in the third digit). Grey dots represent (from the highest mean to the lowest) the Local Maximum Mean portfolio, the Local Minimum L-moment$_4$, the Local Minimum L-moment$_2$, and the Local Maximum L-moment$_3$

Merlin (2006), and Briec, Kerstens and Jokung (2007)) and a set of robust statistics called L-moments. In this framework, the portfolio efficiency is simply evaluated by looking simultaneously for L-variance and L-*kurtosis* contractions, and mean and (positive) L-skewness expansions. We thus approximate the true but unknown Four L-moment Efficient Frontier by a non-parametric portfolio frontier, using an efficiency measure that guarantees a global optimum in a four-dimensional space. In addition, the shortage function approach can adapt itself to any particular multi-moment asset allocation focusing on return-maximization, (L)-skewness-maximization, (L)-variance-minimization, and (L)-*kurtosis*-minimization. Furthermore, dual interpretations are available without imposing any extra simplifying hypotheses (see Briec, Kerstens and Jokung (2007)).

An empirical application provides a four-dimensional representation of the primal

Table 2.2. Statistical Properties of Minimum L-moment$_2$ Portfolio versus the Local Maximum Mean, Local Maximum L-moment$_3$ and Local Minimum L-moment$_4$ Portfolios

	Mean	L-moment$_2$	L-moment$_3$	L-moment$_4$	Sharpe ratio
Minimum L-moment$_2$	.15	.05	−.70	1.30	1.19
Local Maximum Mean	.28	.06	−.90	1.40	2.34
Local Maximum L-moment$_3$	.11	.05	.20	1.70	.71
Local Maximum L-moment$_4$	.17	.06	−1.10	.60	1.29

Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. From the set of Four L-moment Efficient Portfolios with a L-moment2 that is not significantly different from the one of the Minimum L-moment2 portfolio (at a 5% threshold), we select the Local Maximum Mean, Local Maximum L-moment$_3$ and Local Minimum L-moment$_4$ Portfolios (see Figure 2.13). The table presents the first four L-moments (the two first L-moments being annualized and the two last ones being divided by 1,000), and Sharpe ratios of selected local best portfolios.

non-convex four L-moments efficient portfolio frontier and illustrate the computational tractability and relative robustness of the approach. The Efficient Frontier estimation recovered from a sample universe of 162 European stocks shows various rational trade-offs between moments that are coherent with previous approaches (see Athayde and Flôres (2004), Jurczenko, Maillet and Merlin (2006), and Maringer and Parpas (2008)). Nevertheless, comparisons between approximations of traditional CRRA, Quartic and Power-expo Utility functions based on a fourth-order Taylor expansion, computed with C- or L-moments (with decreasing absolute values of sensitivities to moments), show that optimal portfolios mainly lie on the conventional mean-variance efficient frontier in a traditional expected utility setting. In other words, we conclude as Jondeau and Rockinger (2003b and 2006) that higher moments can only matter when investors exhibit major preferences for higher moments and when underlying assets are massively non-Gaussian, that is when we depart a lot from the traditional analysis by Markowitz (1952). However, we also advocate for the use of local optimal higher L-moment efficient portfolios that may have for the investors some advantages over the traditional mean-variance solutions (see also Darolles, Gouriéroux and Jasiak (2008)).

A first natural extension of this study will be to apply the proposed method to non-linear asset payoffs of protected, option-like or alternative investments, where divergences

from Normality are truly large and have specific forms (large pointwise discrepancies, and not only, as in our case study, small diluted differences). A second straightforward potential extension is to be found in a conditional version of our four L-moment asset allocation model. A conditional time-varying modelling of the Four L-moments can be achieved either by using a multivariate GARCH-based model specification for the asset return covariances (see Jondeau and Rockinger (2003a and 2008), and Jondeau, Poon and Rockinger (2007)), or within a robust dynamic autoregressive quantile model (see the CAViaR model by Engle and Manganelli (2004), the QAR by Koenker and Xiao (2006), the CHARN by Martins-Filho and Yao (2006), the DAQ by Gouriéroux and Jasiak (2008), the CARE and EWQR by Taylor (2008a and 2008b)). A third possible extension of our work would be in the development of an asset pricing relation, exploiting the potential improvement of our new notation of higher-order moments, that generalizes the Gini-CAPM (see Shalit and Yitzhaki (1989), Okunev (1990), and Benson, Faff and Pope (2003)) in the four L-moment direction, which can be done for instance by imposing some restrictions on the joint asset return distribution (see Shalit and Yitshaki (2005)), and possibly estimated with Realized (L-)comoments (see Galagedera and Maharaj (2008)). Finally, it would be of great interest to investigate the relative performance of our Four L-moment approach with respect to alternative robust non-parametric (see Ledoit and Wolf (2003 and 2004), Kim and White (2004), and Gouriéroux and Liu (2006)) and parametric (Adcock (2008), Buckley, Saunders and Seco (2008), Mencía and Sentana (2008), Cvitanić, Polimenis and Zapatero (2008)) extended multi-moment asset allocation and pricing models in an international setting (see Guidolin and Timmerman (2008)).

2.6 Appendix

2.6.1 Appendix 1

We briefly present below the main utility functions and their approximations based on the first four moments used in Figure 12.

Let $U(.)$ be a general utility function of the end-of-period wealth W. A Taylor series

expansion allows us to obtain the expected utility (see Jondeau and Rockinger (2006), Jurczenko and Maillet (2006a), and Garlappi and Skoulakis (2008) for precise conditions[39] for the development to be exact or approximative), written as such:

$$\begin{aligned} E\left[U\left(W\right)\right] = U\left(\overline{W}\right) + U^{(1)}\left(\overline{W}\right) E\left[W - \overline{W}\right] + 2^{-1} U^{(2)}\left(\overline{W}\right) E\left[\left(W - \overline{W}\right)^2\right] \\ + \left(3!\right)^{-1} U^{(3)}\left(\overline{W}\right) E\left[\left(W - \overline{W}\right)^3\right] + \left(4!\right)^{-1} U^{(4)}\left(\overline{W}\right) E\left[\left(W - \overline{W}\right)^4\right], \end{aligned} \tag{2.42}$$

where $E(.)$ is the expectation operator, $U^{(n)}$, for $n = [1, ..., 4]$, corresponds to the *n-th* derivative of the utility function with respect to her final wealth denoted W, with W_0 the initial wealth (equal to the unity by simplification[40]), $\overline{W} = E\left(W\right) = \left[1 + E\left(R_p\right)\right]$ the expected final wealth, $R_p = W/W_0 - 1$ the random return on the portfolio held by the investor.

Thus, if we consider a Constant Relative Risk Aversion Utility function (see Mehra and Prescott (1985)) such as (with previous notations):

$$U\left(W\right) = \left(1 - \gamma\right)^{-1} W^{1-\gamma}, \tag{2.43}$$

[39] Even if the Taylor series expansion is a standard and a well established way to express utility functions in terms of moments (see Samuelson (1970)), it is to be noticeable that all moment preferences (first derivatives) only depend upon the first moment of return (alone). This feature is not merely intuitive since an investor targeting a high return portfolio would be more sensitive to extreme events compared to an investor choosing a less profitable portfolio. This claims for a more general utility representation than the usual ones we tested here. In addition, Loistl (1976) argues that such approximations could also be largely biased if some conditions are not respected (see Garlappi and Skoulakis (2008) for details on these conditions).

[40] A well-known criticism of portfolio choice models is the absence of any wealth effect in the analysis (see Quizon, Binswanger and Machina, (1984)); some authors suggest that the relative risk aversion increases with wealth, whilst others conclude with the opposite effect (see Peress, (2004)). Nevertheless, the main approach in the investor preference literature is to scale the initial agent wealth to one. We first thought that such an approach could be very restrictive and, accordingly, we also tested optimal utility-linked choices considering various levels of wealth (as an attempt to intensify the role of the higher moments). Our first results were not, however, conclusive. More precisely, we did not find any clear impact on the moment preferences. Added to the fact that the utility function approximations become more complicated in terms of dependence to the higher moments (due to the presence of various wealth-moment interactions), we, secondly, chose to rescale to one the initial wealth for the sake of simplicity. This renormalization allows us to write simply the expansion and to easily make the link between moments and preferences in the context of series expansion. It ensures also the complete equivalence between the expected utility expressions both in terms of return and terminal wealth - with a simple substitution of W (in this Appendix and in Figure 12) by R_p (in the *corpus* of the text). See Brockett and Golden (1987) and Jondeau and Rockinger (2006) for discussions regarding completely monotone utility functions and independence to wealth levels (see also Footnote 35 in the text).

we obtain using C-moments:

$$\begin{aligned} E[U(W)] = (1-\gamma)^{-1}(m^1)^{1-\gamma} - 2^{-1}\gamma m^2 (m^1)^{-\gamma-1} + (3!)^{-1} \times \\ \gamma(\gamma+1) m^3 (m^1)^{-\gamma-2} - (4!)^{-1}\gamma(\gamma+1)(\gamma+2) m^4 (m^1)^{-\gamma-2}, \end{aligned} \tag{2.44}$$

with $\gamma =]0, 20] \setminus \{1\}$ and where the m^n, for $n = [1, ..., 4]$, are the first four centered moments of the returns.

If we now consider a Quartic Utility function (see Jondeau and Rockinger (2006) and Jurczenko and Maillet (2006a)) as such (with previous notations):

$$U(W) = \beta_0 + \beta_1 W - \beta_2 W^2 + \beta_3 W^3 - \beta_4 W^4, \tag{2.45}$$

we have under the same restrictions:

$$\begin{aligned} E[U(W)] = a_0 + a_1 m^1 - a_2 (m^1)^2 + a_3 (m^1)^3 + a_4 (m^1)^4 + \\ [a_2 + 3a_3 m^1 + 6a_4 (m^1)^2] m^1 + [a_3 + 4a_4 m^1] m^3 + a_4 m^4. \end{aligned} \tag{2.46}$$

where, for $n = [1, ..., 4]$, the m^n being the first four centered moments of the returns and $a_i \in IR$.

If we finally consider a Power-exponential Utility function (see Saha (1993), and Holt and Laury (2002)) as such (with previous notations):

$$U(W) = \alpha^{-1}\left[1 - \exp\left(-\alpha W^{1-\gamma}\right)\right], \tag{2.47}$$

we now have using C-moments:

$$E[U(W)] = \beta_1 + 2^{-1}\beta_2 m^2 + (3!)^{-1}\beta_3 m^3 + (4!)^{-1}\beta_4 m^4, \tag{2.48}$$

where, for $n = [1, ..., 4]$, the m^n being the first four centered moments of the returns and the $\beta_n \in IR$ are such as:

$$\begin{cases} \beta_1 = \alpha^{-1}\left\{1-\exp\left[-\alpha(m^1)^{1-\gamma}\right]\right\} \\ \beta_2 = (1-\gamma)\exp\left[-\alpha(m^1)^{1-\gamma}\right] \times \left[-\gamma(m^1)^{1-\gamma} - \alpha\,(1-\gamma)\,(m^1)^{-2\gamma}\right] \\ \beta_3 = (1-\gamma)^2\exp\left[-\alpha(m^1)^{1-\gamma}\right] \times \left[2\alpha\gamma(m^1)^{-1-2\gamma} + \alpha^2\,(1-\gamma)\,(m^1)^{-3\gamma}\right. \\ \qquad\qquad \left. -\gamma(m^1)^{-\gamma} + \alpha(m^1)^{-1-2\gamma}\right] \\ \beta_4 = (1-\gamma)^2\exp\left[-\alpha(m^1)^{1-\gamma}\right] \times \\ \qquad \left[2\alpha\gamma\,(2\gamma-1)\,(m^1)^{-2-2\gamma} - 5\alpha^2\gamma\,(1-\gamma)\,(m^1)^{-1-3\gamma} + \gamma^2(m^1)^{-\gamma-1}\right. \\ \qquad \left. \alpha\,(-\gamma^2-\gamma+1)\,(m^1)^{-2\gamma} - \alpha^3\,(1-\gamma)^2\,(m^1)^{-4\gamma} - \alpha^2\,(1-\gamma)\,(m^1)^{1-3\gamma}\right], \end{cases}$$

with $\alpha \in IR \setminus \{0\}$, $\gamma \in IR \setminus \{1\}$, $\alpha\,(1-\gamma) > 0$, where α and $(1-\gamma)$ respectively govern the Relative and Absolute Risk Aversions. ■

2.6.2 Appendix 2

We hereafter recall the propositions given in the *corpus* of the article regarding the optimization programs and present their proofs.

Proposition 1. *For every* $\mathbf{g} \in (IR_+ \times IR_- \times IR_+ \times IR_-) \setminus \{0\}$ *and every* $\mathbf{w} \in \mathfrak{F}$*, there exists a unique element* δ^**,* $\delta^* \in IR_+$*, such that:*

$$S_{\mathbf{g}}(\mathbf{w}) = \lambda_{\mathbf{w}} + \delta^*\mathbf{g}. \tag{2.34}$$

Proof of Proposition 1.

First, we recall that the set of the feasible portfolios $\mathfrak{A}$ can be expressed as follows:

$$\mathfrak{A} = \left\{\mathbf{w} \in IR^N : \mathbf{w}'\mathbf{1} = 1 \text{ and } \mathbf{w} \geq \mathbf{0}\right\}, \tag{2.26}$$

where $\mathbf{w}'$ is the $(1 \times N)$ transposed vector of the investor's holdings in the various risky assets, $\mathbf{1}$ is the $(N \times 1)$ unitary vector and $\mathbf{0}$ is the $(N \times 1)$ null vector. Also recall that the set of the feasible portfolios in the four L-moment space in a free disposal representation, denoted $\mathfrak{F}$ (see Briec, Kerstens and Lesourd (2004), and Briec, Kerstens and Jokung (2007)) writes:

$$\mathfrak{F} = \{\lambda_{\mathbf{w}} : \mathbf{w} \in \mathfrak{A}\} + \left[(-IR_+) \times IR_+ \times (-IR_+) \times IR_+\right], \tag{2.27}$$

where $\lambda_{\mathbf{w}}$ is the (4×1) vector of the first four L-moments of the portfolio return $R_{\mathbf{w}}$, *i.e.*:

$$\lambda_{\mathbf{w}} = \left[\lambda_1 (R_{\mathbf{w}}) ; \ \lambda_2 (R_{\mathbf{w}}) ; \ \lambda_3 (R_{\mathbf{w}}) ; \ \lambda_4 (R_{\mathbf{w}})\right]' .$$

Secondly, we recall the definition of the (generic) partial order relation, denoted by $\succeq$, on IR^4, that is for any $(\lambda, \widetilde{\lambda}) \in (IR^4)^2$:

$$\lambda \ \succeq \ \widetilde{\lambda} \iff [\lambda_1 \geq \widetilde{\lambda}_1, \ \lambda_2 \leq \widetilde{\lambda}_2, \ \lambda_3 \geq \widetilde{\lambda}_3, \ \lambda_4 \leq \widetilde{\lambda}_4], \tag{2.28}$$

altogether with a strict relation, denoted by $\succ$, as such:

$$\lambda \ \succ \ \widetilde{\lambda} \iff [\lambda_1 > \widetilde{\lambda}_1, \ \lambda_2 < \widetilde{\lambda}_2, \ \lambda_3 > \widetilde{\lambda}_3, \ \lambda_4 < \widetilde{\lambda}_4]. \tag{2.29}$$

Thirdly, for every $\mathbf{w} \in \mathfrak{A}$ and for every $\mathbf{g} \in (IR_+ \times IR_- \times IR_+ \times IR_-) \setminus \{0\}$, let us define the set $\mathfrak{D}_{\mathbf{w},\mathbf{g}}$ of admissible distances from the portfolio $\mathbf{w}$ to the weakly efficient frontier in the direction $\mathbf{g}$, such as:

$$\mathfrak{D}_{\mathbf{w},\mathbf{g}} = \{\delta \in IR_+ : \ \ (\lambda_{\mathbf{w}} + \delta \mathbf{g}) \in \mathfrak{F}\} . \tag{2.49}$$

Let now the mapping $\Lambda(.)$ from $\mathfrak{A}$ to IR^4 defined by, for every $\mathbf{w} \in \mathfrak{A}$:

$$\Lambda(\mathbf{w}) = \lambda_{\mathbf{w}}. \tag{2.50}$$

The mapping $\Lambda(.)$ is continuous since all coordinate functions are polynomial functions. Moreover, since $\mathfrak{A}$ is a compact set, then $\Lambda(\mathfrak{A})$ is also a compact set (see Theorem 4.14, p.89, in Rudin (1976)). We thus have:

$$\mathfrak{F} = \Lambda(\mathfrak{A}) + [(-IR_+) \times IR_+ \times (-IR_+ \times IR_+)] . \tag{2.51}$$

So there exists $\overline{\lambda}_{\mathbf{w}} \in IR^4$ such that for every $\lambda_{\mathbf{w}} \in \mathfrak{F}$, $\overline{\lambda}_{\mathbf{w}} \succeq \lambda_{\mathbf{w}}$. This implies that $\mathfrak{D}_{\mathbf{w},\,\mathbf{g}}$ is bounded.

The set $\mathfrak{F}$ is the sum of a compact set and a closed set, so it is a closed set. Moreover,

the mapping $\delta \mapsto (\lambda_{\mathbf{w}} + \delta \mathbf{g})$ is continuous on IR_+, and then $\mathfrak{D}_{\mathbf{w},\mathbf{g}}$ is closed. We deduce that $\mathfrak{D}_{\mathbf{w},\,\mathbf{g}}$ is also a compact. Then, there exists $\delta^* \geq 0$ such that $S_{\mathbf{g}}(\mathbf{w}) = (\lambda_{\mathbf{w}} + \delta^* \mathbf{g})$.

At this point, we note, however, that the mapping $\Lambda(.)$ is not necessarily bijective (depending on the considered set of portfolio characteristics), meaning that several different portfolios may potentially be located on the same point of the efficient frontier. In order to avoid this uncertainty, we can also think in terms of equivalence classes of portfolios, formally defining the equivalence relation on the set of portfolios by:

$$\mathbf{w} \sim \breve{\mathbf{w}} \iff \Lambda(\mathbf{w}) = \Lambda(\breve{\mathbf{w}}), \tag{2.52}$$

for every $\mathbf{w}$ and $\breve{\mathbf{w}}$ in $\mathfrak{A}$.

Then, the equivalence class associated to a portfolio $\mathbf{w}$, and denoted $\dot{\mathbf{w}}$, is:

$$\dot{\mathbf{w}} = \{\breve{\mathbf{w}} \in \mathfrak{A} \mid \mathbf{w} \sim \breve{\mathbf{w}}\}, \tag{2.53}$$

where the sign $\sim$ stands for the equivalence relation expressed in terms of utility for the investor.

We also remark that the fact that a set of L-moments determine a unique density function (on the contrary of C-moments, see the so-called Hamburger problem - *Cf.* Hamburger (1920), Jondeau and Rockinger (2003a), and Jurczenko and Maillet (2006a)) allows us to make the link between the weights of a portfolio, the density of returns on it, its L-moments and the utility function of the investor (provided that the density function can be perfectly described by the first four L-moments, which is the case for a large family of four parameter densities). For the sake of simplicity, however, we considered in the *corpus* of the text the term portfolio weights, even if, strictly speaking, we should have referred to equivalence classes of these portfolio weights. ■

Proposition 2. *If* $(\mathbf{w}^*, \delta^*) \in (\mathfrak{A} \times IR_+)$ *is a local solution of the following non-linear optimization program* $\mathfrak{P}_{\mathbf{w}_p,\,\mathbf{g}}$.

$$\underset{(\mathbf{w},\delta)\in(\mathfrak{A}\times\, IR_+)}{\text{Max}} \Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}), \tag{2.40}$$

it is then a global solution.

Proof of Proposition 2.

Let us denote:

$$\mathfrak{J} = \left\{(\mathbf{w},\delta) \in (\mathfrak{A} \times\, IR_+) :\ \lambda_{\mathbf{w}} \ \succeq\ (\lambda_{\mathbf{w}_p} + \delta\mathbf{g})\right\}, \tag{2.54}$$

which is the hypograph of the mapping $\Phi_{\mathbf{w}_p,\,\mathbf{g}}(.)$ with the order relation, denoted $\succeq$ and defined above.

We then have:

$$\Phi_{\mathbf{w}_p,\,\mathbf{g}}(\mathbf{w}) = \underset{\delta\in IR_+}{\text{Sup}}\ \{\delta : (\mathbf{w},\,\delta) \in \mathfrak{J}\}. \tag{2.55}$$

Assume that the couple $(\mathbf{w}_1,\,\delta_1)$ constitutes a local maximum, but is not a global one. In that case, there exists a couple $(\mathbf{w}_2, \delta_2) \in \mathfrak{J}$ such that:

$$\delta_2 > \delta_1. \tag{2.56}$$

This implies that for all $\delta \in [\delta_1,\,\delta_2]$, $(\mathbf{w}_2,\,\delta) \in \mathfrak{J}$. Therefore, a neighborhood $\mathcal{N}[(\mathbf{w}_1,\ \delta_1),\ \varepsilon]$ where $\varepsilon > 0$, such that $\delta_1 \geq \delta$ for all $(\mathbf{w},\,\delta) \in \mathcal{N}[(\mathbf{w}_1,\,\delta_1),\varepsilon]$ does not exist. Consequently, if $(\mathbf{w}^*,\,\delta^*)$ is a local maximum, then it is also a global maximum. ■

2.6.3 Appendix 3

We briefly present herein the data mining technique we applied for grouping together optimal portfolios as illustrated in Figure 2.10 (and in Figure 2.14 below). Self-Organizing Maps (SOM or Kohonen Maps) are a clustering method with their roots in Artificial Neural Networks. SOM can be used at the same time both to reduce the amount of relevant data by clustering, and for projecting the data nonlinearly onto a lower dimensional display. Due to its unsupervised learning and topology preserving proprieties, the SOM algorithm

has proven to be especially suitable in visual analysis of high dimensional sets. They have already been applied in various fields in general, and in finance in particular, for clustering elements sharing some similarities. With no pretension of exhaustivity, some examples of SOM' financial applications are to be found in Deboeck and Kohonen (1998), Resta (2001), Maillet and Rousset (2003), Das and Das (2004), Moreno, Marco and Olmeda (2006), and Ben Omrane and de Bodt (2007). For further details on this data-mining technique, see Kohonen (2000) and Guinot, Maillet and Rousset (2006).

A direct characterization of (65,253) optimal portfolios based on the four L-moment computations is not an easy task. It is very natural to try to reduce the dimension of the problem by considering an exploratory algorithm such as SOM. Our goal is here to group optimal portfolios according to their first four L-moments and to visualize their various locations relative to the efficient frontier. SOM thus enable us to create homogeneous clusters as well as virtual portfolios representative of each cluster. In the following, we will consider the $(65,253 \times 4)$ matrix $\mathbf{X}$, where each line $\mathbf{x}$ corresponds to the estimated first four L-moments of any optimal portfolio.

In order to define the network, the number of neurons (*i.e.* the number of clusters) as well as the shape of the grid (a string or a lattice generally) have to be specified. Let us consider K the number of neurons (units or "code vectors") and $I = [1, 2, ..., K]$ the total set of neurons.

The network state at the iteration j is given by:

$$\mathbf{V}(j) = [\mathbf{V}_1(j), \mathbf{V}_2(j), ..., \mathbf{V}_K(j)], \tag{2.57}$$

where $\mathbf{V}_i(j)$ is the weight vector of unit i, its size being equal to the dimension of the data to be classified. In our case, we classify portfolios according to their first four L-moments, each weight vector is thus a 4-dimensional vector.

The network is first randomly initialized (iteration 1) from the input set (with previous

notations):

$$\begin{aligned}\mathbf{V}(1) &= [\mathbf{V}_1(1), \mathbf{V}_2(1), ..., \mathbf{V}_K(1)] \\ &= \left\{ \begin{bmatrix} \hat{\lambda}_1^1(1) \\ \hat{\lambda}_2^1(1) \\ \hat{\lambda}_3^1(1) \\ \hat{\lambda}_4^1(1) \end{bmatrix}, \begin{bmatrix} \hat{\lambda}_1^2(1) \\ \hat{\lambda}_2^2(1) \\ \hat{\lambda}_3^2(1) \\ \hat{\lambda}_4^2(1) \end{bmatrix}, \ ... \ , \begin{bmatrix} \hat{\lambda}_1^K(1) \\ \hat{\lambda}_2^K(1) \\ \hat{\lambda}_3^K(1) \\ \hat{\lambda}_4^K(1) \end{bmatrix} \right\}\end{aligned} \tag{2.58}$$

where λ_k^i, with $i \in I$ and $k = [1, ..., 4]$, corresponds to the empirical $k-th$ L-moment of a portfolio randomly drawn from $\mathbf{X}$.

The SOM algorithm is then recursively defined by the following iterations.

1. Draw randomly another observation $\mathbf{x}$ for the set of optimal portfolio L-moments.

2. Find the associated winning unit $i_w(\mathbf{x}, \mathbf{V})$ also called the Best Matching Unit (noted BMU), that is the unit whose weight $\mathbf{V}_{i_w(\mathbf{x},\mathbf{V})}$ is the closest to input $\mathbf{x}$ (with previous notations):

$$BMU = i_w[\mathbf{x}(j+1), \mathbf{V}(j)] = \underset{(\mathbf{V}_i,\, i) \in (IR^4 \times I)}{\text{ArgMin}} \{\|\mathbf{x}(j+1) - \mathbf{V}_i(j)\|\}, \tag{2.59}$$

where $\|.\|$ is the Euclidian norm.

3. Once the BMU is found, the weight vectors of the SOM are updated so that the BMU and the activated neighbors are moved closer to the input vector. The SOM update rule is, $\forall i \in I$ (with previous notations):

$$\mathbf{V}_i(j+1) = \mathbf{V}_i(j) - \tau(j)\, \mathbb{k}_{BMU,i}(j)\, [\mathbf{V}_i(j) - \mathbf{x}(j+1)], \tag{2.60}$$

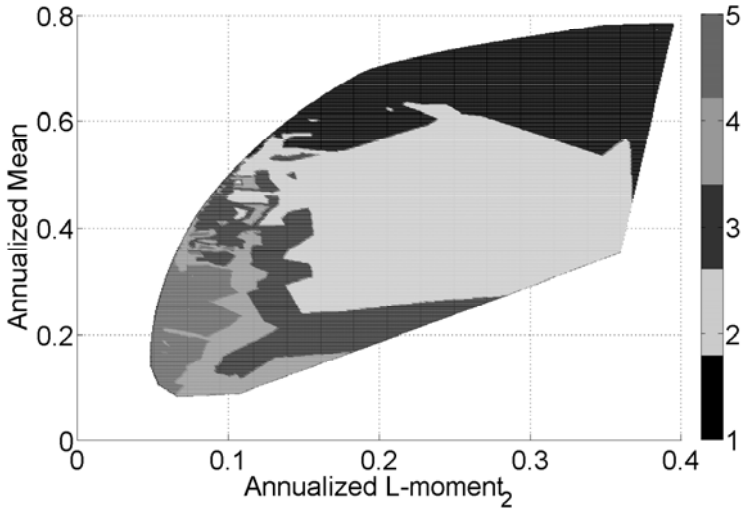
where $\tau(.)$ is the learning rate function, which is a $]0, 1[$-valued decreasing function. The function $\mathbb{k}_{BMU,i}(.)$ is the neighborhood (Gaussian) kernel around the winner unit BMU, that also decreases with iterations.

Following Guinot, Maillet and Rousset (2006), we choose to apply a robust version of the Kohonen algorithm named "Robust Maps". Through an intensive use of bootstrap, this enables us to overcome the potential convergence problem of the general algorithm. Also for practical considerations, we choose to classify all optimal portfolios using their

empirical first four L-moments, and to group and project them onto an arbitrary string compound of five neurons. Thus, the overall algorithm leads to the creation of five homogeneous clusters of four L-moment efficient portfolios.

Figure 2.10 in the text is based on the classification of raw estimated L-moments of optimal portfolios. In order to make the picture more clear, and also to mitigate the problem of L-moment differences in scale (which is, however, less critical than for C-moments, since L-moments are all rescaled differences of returns and not of the magnitude of returns at various powers), we also run the algorithm on quantiles (ranking) of the L-moments of optimal portfolios (based on their empirical cumulative distributions). The aim is here to cluster the highest values of the population L-moment of efficient portfolios together, taking into account all first L-moments at the same time. The resulting classification (see Figure 2.14 below) based on estimated quantiles of L-moments slightly differs from the first classification based on raw L-moments (see Figure 2.10 in the text). The clusters seem to be more homogeneous (more concentrated), with a clear breakdown operated by the two first L-moments (along the traditional efficient frontier). Thus, the clusters are sequenced by descending order in terms of the first two L-moments: investors willing large mean returns (and accepting large related second L-moments) should choose a portfolio in cluster 1, whilst a more risk averse agent should pick up a portfolio within cluster 5. Focusing now on higher L-moments, it appears that the first cluster groups together portfolios with a high L-moment$_4$ (and L-moment$_3$ above zero), whilst the second cluster (in the center of the *parabola*) exhibits a high average L-moment$_3$. Lowest L-moment$_4$ portfolios are grouped in the fourth cluster, whereas those with the lowest (negative) third L-moments mostly belong to the fifth cluster. Finally, the third cluster corresponds to portfolios with medium L-moment$_4$ values. Overall, we can see from this clusterization, based on quantiles of L-moments (rather than on L-moments themselves), that regions in the projected efficient plane may be easily explained by the various L-moment characteristics of the optimal primal efficient portfolios.

Figure 2.14. Clusters of the Four L-moment Optimal Portfolios according to their L-moment Rankings in the L1-L2 Plane.



Source: *Bloomberg*, weekly net asset Values (06/2001-06/2006) in EUR; computations by the authors. The colours represent the clusters of the Four L-moment Efficient Portfolios according to their L-moment ranking similarity. This figure is obtained using a Self-Organizing Maps' classification algorithm on a string of five cells, applied on vectors of L-moment empirical quantiles related to optimal portfolios (see Kohonen (2000)).

Chapter 3

Outliers Detection, Correction of Financial Time-series Anomalies and Distributional Timing for Robust Efficient Higher-order Moment Asset Allocations

3.1 Introduction

The estimation of moments is a key issue in financial modelling. Under the Gaussian hypothesis of asset return distribution, numbers of models were first based on expected returns and standard deviations. With the recognition that asset returns do not comply with such a paradigm, some developments regarding expansions with higher-order moments, namely skew and *kurtosis*, have been proposed over the last twenty years in various fields such as option pricing (*Cf.* Corrado and Su, 1996), asset pricing (*Cf.* Harvey and Siddique, 2000) and asset allocation (*Cf.* Athayde and Flôres, 2002). However, as pointed out by Kalymon (1971), Jorion (1986) and Michaud (1989), for instance, the estimation risk should be taken into account in financial models. More specifically, as

shown by Rosenberg and Houglet (1974), the higher the order of the estimated moment required, the more the estimation is likely to be biased. Indeed, empirical moments are highly subject to bias due to the potential influence of so-called "outliers". In a broad definition, outliers are considered to be some realizations that are not likely to happen regarding a supposed distribution (*Cf.* Barnett and Lewis, 1978). This first general acceptance led to three more formal definitions in finance. The first one consists of defining an outlier according to a chosen distribution as well as a threshold upon which the realization is supposed to be "aberrant" (*Cf.* Johansen and Sornette, 2001; Gonzalo and Olmo, 2004). The second one explicitly considers a time-series structure of the financial variable before evaluating realizations that are not likely to happen (*Cf.* Carnero *et al.*, 2007). Thirdly, outliers may also be defined as consequences of economic, political or financial events that have been observed in financial time-series but that are very unlikely to happen again (*Cf.* Franses and van Dijk, 2000). As the Gaussian distribution remains the standard when modelling financial returns, "abnormal" returns are often considered as outliers. This leads to an ambiguity due to the fact that "abnormal" could stand for non-Gaussian but also for not common. This highlights the importance of distinguishing extreme from aberrant returns when dealing with outliers in finance, since the confusion is still common (*Cf.* Gonzalo and Olmo, 2004). As shown by Johansen and Sornette (2001), only a minor part of these extreme returns (the most extreme) should be considered as outliers.

A number of studies emphasize the impact of outliers on the moments of asset return distributions and the resulting bias in optimal portfolio determination. For instance, Best and Grauer (1992) provide a complete study of the sensitivity of the mean-variance efficient frontier to abnormal returns. Due to the high impact of these errors on efficient portfolios, several methods to limit the effect of these anomalies have been proposed. First, we can use robust approaches (*Cf.* Hampel *et al.*, 2005) to estimate efficient portfolios. For instance, we can apply resampling methods, such as a bootstrap procedure in Michaud (1998), to reduce the impact of outliers on optimized portfolios, or we can directly con-

sider robust statistics instead of conventional variance, as in the Mean-Gini Capital Asset Pricing proposed by Shalit and Yitshaki (1989). Secondly, we can also apply an outlier detection model and correction method as a pre-processing method before running the traditional approach applied on the denoized series of returns. Chen and Liu (1993), for instance, propose modelling returns with an AutoRegressive Conditional Heteroskedasticity (ARCH) model and develop a method to correct the detected outliers. Their seminal work in the field leads to several developments and is still a standard when detecting and correcting outliers in a financial return database (*Cf.*, among others, Franses and Ghijsels, 1999; Franses and van Dijk, 1999; Charles, 2004; Doornik and Ooms, 2005; Charles and Darné, 2005 and 2006).

However, as pointed out by Grané and Veiga (2009), this main stream approach for correcting outliers has to be followed with caution. On one hand, neglecting the existence of (some) outliers during the estimation phase of the detection methodology may end up with biased parameters (*Cf.* Fox, 1972; van Dijk *et al.*, 1999). Consequently, a potential issue about the methodology is that outliers are defined according to statistics based on non-accurate GARCH parameters. On the other hand, the strong empirical evidence that asset returns do not follow a Gaussian distribution, questions the hypothesis of a Normal GARCH model followed by the returns and leads some authors to incorporate higher-order moments of return distributions in the field of Efficient Portfolio Selection (*Cf.* Lai, 1991; Chunhachinda *et al.*, 1997; Jondeau and Rockinger, 2006; and Jurczenko *et al.*, 2006). However, some authors at the same time (for instance, Kim and White, 2004) argue that the consideration of higher-order moments is useless, since a large part of these measures comes from outliers (*Cf.* Jondeau and Rockinger, 2009, for a study of the effect of shocks on higher-order moments).

The goal of this paper is thus two-fold. We propose to modify the standard GARCH-based outlier detection model proposed by Franses and Ghijsels (1999) and Franses and van Dijk (1999), first dealing with critics of the non-robustness of the Quasi-Maximum Likelihood estimation method and second, with the introduction of an extended Artificial

Neural Network GARCH model (denoted ANN-GARCH, *Cf.* Donaldson and Kamstra, 1997; Miazhynskaia *et al.*, 2006; Roh, 2007; Medeiros *et al.*, 2008 and Bildirici and Ersin, 2009) that accommodates more efficiently the empirical peculiarities of asset returns regarding their non-normality and non-linearity features. Even if specific GARCH models (for instance GARCH with Student, NIG or GEV distributions) may be more appropriate for the study of outliers, we choose in the following to keep as the benchmark the simplest volatility model since, first, Gaussian GARCH models are still widely used (*Cf.* Hansen and Lunde, 2005) and, secondly, because we expect that the Artificial Neural Network better adjusts the true empirical distribution and explains the non-linear part of the Gaussian GARCH residual. Finally, keeping the standard model as a benchmark allows us to compare our results with those of Franses and Ghijsels (1999) and Charles and Darné (2005). We also consider in the following the impact of outliers on the higher-order moment asset allocation model proposed by Jurczenko *et al.* (2006), as well as on its robust version by Jurczenko *et al.* (2008). As Jondeau and Rockinger (2009) did, we show the importance of considering robust statistics when using higher-order moment-based models applied to some distibutional strategies, evaluated in the Expected Utility and the Cumulative Prospect Theory frameworks.

This paper is organized as follows. Section 3.2 is dedicated to a brief presentation of the ANN-GARCH model (*Cf.* Bildirici and Ersin, 2009), and to the outlier detection-correction model proposed by Franses and Ghijsels (1999) and generalized by Charles and Darné (2005). It also shows how the Artificial Neural Network GARCH may be incorporated to improve the outlier detection procedure. In section 3.3, we present the two versions of higher-order moment asset allocation models (traditional and robust) proposed by Jurczenko *et al.* (2006 and 2008). Section 3.4 is devoted to empirical illustrations which apply the outlier detection procedure to a CAC40 daily stock market dataset. Finally, we compare the feature of some notorious Efficient Portfolios provided by four moment asset allocation models (robust or not) when the data are raw or corrected from outliers. The last section concludes.

3.2 Detecting and Correcting the Outliers

In the following section, we shall start by presenting the basics of GARCH modelling as well as its extended ANN-GARCH version. We then show how the related implied time-series structures could be used to detect outliers. We then present the various ways to correct the outliers for cleaning the market data and show in more detail how to correct outliers when using both Gaussian-GARCH and ANN-GARCH time-series structures.

3.2.1 From GARCH to ANN-GARCH Modelling

Literature dealing with volatility models has been extensive since the seminal work of Engle (1982) regarding the AutoRegressive Conditional Heteroskedasticity (ARCH) model of volatility that has the ability to reflect the observed volatility clusters (*Cf.* Mandelbrot, 1963). Bollerslev (1986) proposed a generalization through the ARCH (named GARCH), with a conditional variance of the innovation term in the return equation being assumed to depend linearly on past volatilities as well as on past shocks. These models represent a certain type of conditional heteroskedasticity characterized by successive periods of high and low volatility in the history of a time-series. The original GARCH model imposes symmetry on the response of the variance to past shocks or "news", where the volatility depends only on the size and not on the sign of the shock.

A GARCH(p, q) model takes the general strong form (*Cf.* Bollerslev, 1986, and Drost and Nijman, 1993):

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ h_t = \alpha_0 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \end{cases} \tag{3.1}$$

where ε_t, for $t = [1, \ldots, T]$, is a sequence of standardized innovations (returns in our case), η_t is assumed to be distributed as standard normal, with sufficient conditions for conditional variance positivity and unconditional variance existence being that $\alpha_i \geq 0$ for $i = [0, \ldots, q]$, with $q \in \mathbb{N}$, $\beta_i \geq 0$ for $i = [1, \ldots, p]$, with $p \in \mathbb{N}^*$, and $\sum_{i=1}^{m} \delta_i < 1$, where $\delta_i = (\alpha_i + \beta_i)$, for $i = [0, \ldots, m]$ and $m = \max(p, q)$ and with *per* convention: $\alpha_i = 0$ for $i > q$ and $\beta_0 = \beta_i = 0$ for $i > p$. The GARCH family models found some impor-

tant applications in finance where volatility plays a role. However, two major drawbacks mitigate the explanation power of these approaches. First, GARCH models are generally estimated with a Quasi Maximum Likelihood procedure which may lead to estimations of biased parameters (*Cf.* Zumbach, 2000). Secondly, these methods become inappropriate if errors are assumed to be "strongly" non Gaussian. The fact that residual is highly non-normal (*Cf.* Jondeau and Rockinger, 2003) led to the introduction of fat-tailed distribution related versions. Various alternative leptokurtic distributions such as the t-Student (*Cf.* Bollerslev, 1987), the General Error Distribution (*Cf.* Nelson, 1991) or the Normal Inverse Gaussian (*Cf.* Barndorff-Nielsen, 1997; Anderson, 2001) have been proposed to improve the original Gaussian GARCH model's performance, but unfortunately, the study of the related GARCH devolatilized residuals show that they are still significantly non-Gaussian.

Following the model proposed by Donaldson and Kamstra (1997), we hereafter introduce a hybrid model that combines GARCH and an Artificial Neural Network model in an outlier correction framework. It consists of a traditional GARCH model enhanced with a standard MultiLayer Perceptron (denoted hereafter as MLP). The goal of the latter is to explain some of the non-linear parts of the GARCH residual. The ANN-GARCH(p, q) model takes the general form (with previous notations):

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ h_t = \alpha_0 + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \mathfrak{M}\left(\mathbf{I}_{t-1}\right). \end{cases} \tag{3.2}$$

where $\mathbf{I}_{t-1} = [h_{t-1}, \ldots, h_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}]$ and $\mathfrak{M}(.)$ stand, respectively, for the inputs of the MLP and the general non-linear function corresponding to the MLP.

The main reason for adopting this specification (Equation 3.2) is to provide more flexibility in the approximation of mean and conditional volatility, while still taking advantage of the simple and parsimonious specification that a GARCH model offers. Based on the initial approach, the methodology presented below can also be applied to a wider class of ANN-GARCH models, possibly including asymmetric GARCH specifications (*Cf.* Bildirici and Ersin, 2009). In our case, and following the same architecture as Donaldon

and Kamstra (1997), the chosen MLP consists of a three layer network (the input layer, a hidden neuron layer and the output layer). Every neuron on a layer is fully connected to all neurons on the next layer (see Figure 3.1 for a simple illustration[1]).

More formally, we denote by $\mathbf{\Lambda}_{i,j}^{l}$ the weight of the connection between the neuron i of layer l and the neuron j of layer $(l+1)$. The MLP takes as input the vector of information $\mathbf{I}_{t-1}$, propagates this vector's value through hidden layers, and computes the activation of the output neuron (*via* a traditional logistic function). The function corresponding to the MLP, that we note $\mathfrak{M}(\mathbf{I}_{t-1})$, thus depends upon $\mathbf{I}_{t-1}$ and also $\mathbf{\Lambda}$, the weight of each connection. If the input layer has $(p+q)$ neurons and the output layer has a single neuron, then the MLP will appear as a non-linear function from $\mathbb{R}^{p+q}$ into $\mathbb{R}$. The function $\mathfrak{M}(.)$ is non-linear and differentiable. In our case ($p+q$ input neurons, a hidden layer with k neurons and a unique output layer), the function $\mathfrak{M}(.)$ can be defined as such:

$$\mathfrak{M}(\mathbf{I}_{t-1}) = \sum_{j=1}^{k} \left[\Lambda_{j,1}^{2} \Phi \left(\sum_{i=1}^{p+q} I_{t-1,i} \; \Lambda_{i,j}^{1} \right) \right]^{-1}, \tag{3.3}$$

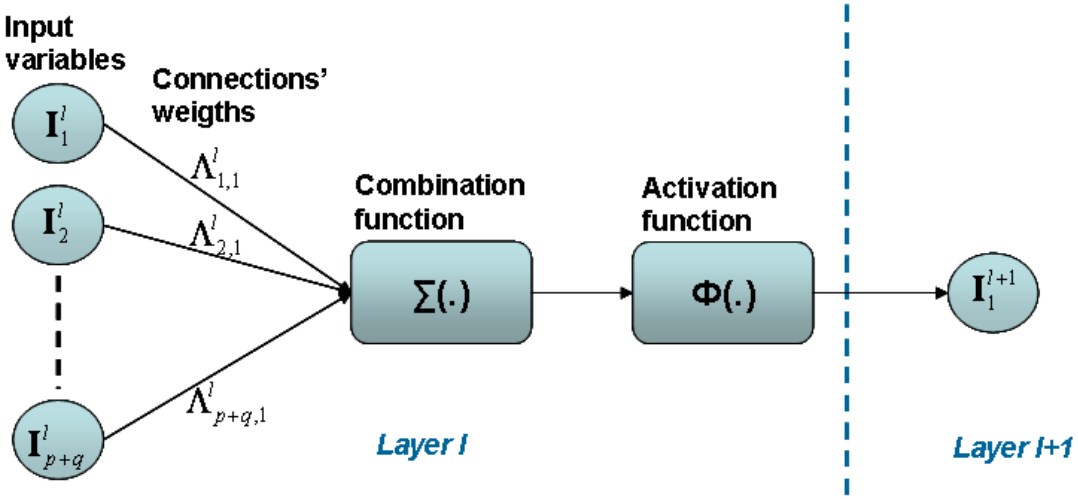
where k is the total number of neurons in the hidden layer, $I_{t-1,i}$ is the i^{th} element of the input vector $\mathbf{I}_{t-1}$, $\Lambda_{i,j}^{1}$ is the weight of the connection between the i^{th} neuron of the input layer and the j^{th} neuron in the hidden layer, $\Lambda_{j,1}^{2}$ is the weight of the connection between the j^{th} neuron in the hidden layer and the output neuron and with $\Phi(.)$ being the non-linear exponential activation function defined as $\Phi(.) = [1 + \exp(.)]^{-1}$.

Regarding the parameter estimation of the ANN-GARCH model, we proceed with a two-step estimation. First, the GARCH model is estimated by using the robust method proposed by Muler and Yohai (2008), called the Bounded Maximum Likelihood method (*Cf.* Appendix 3.6.1), since parameter estimates of GARCH models may be biased if we had followed a standard Quasi-Maximum Likelihood approach (*Cf.* Carnero *et al.*, 2007)[2].

[1] An extensive presentation of MLP can be found in Bishop (1995).

[2] For the sake of security, we also performed a two-pass QML-GARCH parameter estimation. From the first estimation, we detect all significant outliers and re-estimated for a second time the GARCH parameters, but with the outliers removed. The aim of the second estimation round is to mitigate the effect of the most severe outliers on the parameters and can be seen as an estimation on a safe and robust clean data set, when a stretch of the true ouliers are taken to be missing (*Cf.* Grossi, 2004 and Riani, 2004 for extended forward search methods applied to outliers). We then compared the test statistics obtained

Figure 3.1. Simplified Representation of a Typical MultiLayer Perceptron.



Representation of the structure of two successive layers in an MLP. For the sake of simplicity, only one output neuron is represented here.

Then, the unanticipated conditional volatility is deduced as: $\nu_t = \varepsilon_t^2 - h_t$. Secondly, the network learning process is performed, explaining furthermore the unexplained volatility. Knowing both the input vector $\mathbf{I}_{t-1}$ and the target output ν_t, the adaptation of MLP's parameters is made by using the retropropagation algorithm (*Cf.* Bishop, 1995). The structure of the MLP (defined here by k, the number of neurons in the hidden layer) is designed according to a cross-validation technique (*Cf.* Lendasse *et al.*, 2003). In this scheme, the learning process consists of minimizing the quadratic error QE_t between the predicted and the observed values, for $t = [1, \ldots, T]$, such as (with previous notations):

$$QE_t = \sum_{i=1}^{t} \left[\nu_i - \mathfrak{M}\left(\mathbf{I}_{i-1}\right)\right]^2, \tag{3.4}$$

where, for each date t, ν_t and $\mathbf{I}_{t-1}$ stand respectively for the set of outputs (corresponding to the series of unexpected conditional volatilities in our case) and the collection of time-varying input vectors.

in the first run and with their values obtained with the second run parameter estimations. As a result, whilst minor differences were observed in the two sets of test statistics, no large discrepancies were found either in the dates where outliers appear, or in their amplitudes. Moreover, the results obtained by the traditional QML method, the two-run QML and the BQML are very close for all stocks in our sample.

3.2.2 GARCH-related Outlier Detection Models

Based on the Franses and Ghijsels (1999) standard Gaussian GARCH outlier detection model, we adapt the procedure to an ANN-GARCH volatility modelling. Thus, when we consider the return series ε_t, a GARCH(1,1) model is written as (with previous notations):

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 \varepsilon_{t-1}^2. \end{cases} \tag{3.5}$$

The conditions $\alpha_1 + \beta_1 < 1$ and $(\alpha_1 + \beta_1)^2 + \alpha_1^2 (\kappa_\varepsilon - 1) < 1$, with κ_ε the *kurtosis* of innovations, guarantee that the return series has a finite variance and a finite fourth-order moment (*Cf.* He and Teräsvirta, 1999).

The GARCH(1,1) model can be written as an ARMA(1,1) model for the squared returns ε_t^2 (see Bollerslev, 1986):

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \tag{3.6}$$

where $\nu_t = \varepsilon_t^2 - h_t$.

As shown by Franses and Ghijsels (1999), we can exploit the analogy between a GARCH and an ARMA model to adapt the method of Chen and Liu (1993) for detecting both Additive Outliers (denoted hereafter as AO for short) and Innovative Outliers (IO) in an extended GARCH framework. An AO is defined as an exogenous shock that directly affects the series and only its level of observation at a given date. On the contrary, an IO is possibly generated by an endogenous change in the series, and affects all observations after a certain date of arrival through the memory of the process. In the following, we shall start by supposing that there exists a unique outlier in the observed time-series. Later on, we will suppose that the series may contain more than one unique outlier, and the global outlier detection process will simply consist of applying sequentially the very same procedure as for the detection of the unique outlier.

For making clear the distinction between AO and IO in a GARCH framework, we first

need to express ν_t as depending on the past innovations. From Equation (3.5) we have:

$$\begin{aligned} \nu_t &= \varepsilon_t^2 - \alpha_0 - \beta_1 h_{t-1} - \alpha_1 \varepsilon_{t-1}^2 \\ &= -\alpha_0 \left(1 + \beta_1\right)^{-1} + \varepsilon_t^2 - \alpha_1 \varepsilon_{t-1}^2 - \alpha_1 \varepsilon_{t-2}^2 - \beta_1^2 h_{t-2}. \end{aligned} \tag{3.7}$$

which leads to express ν_t in terms of past residuals such as:

$$\nu_t = -\alpha_0 \left(1 - \beta_1\right)^{-1} + \sum_{k=0}^{t-1} \left(\pi_k \varepsilon_{t-k}^2\right), \tag{3.8}$$

where $\pi_k = \left[-\alpha_1 \beta_1^{k-1}\right]^{\mathbb{1}_{\{k>0\}}}$, $k \in \mathbb{N}^+$ and $\mathbb{1}_{\{.\}}$ is the indicator function on the set $\mathbb{N}^*$.

Let us now define an outlier in terms of the squares of the observed time-series. More precisely, we suppose that instead of the true series ε_t, we observe the series e_t that is defined by:

$$e_t^2 = \varepsilon_t^2 + \omega \left[\mathbb{1}_{\{t=\tau\}} + \alpha_1 \left(\beta_1 - \alpha_1\right)^{t-\tau-1} \mathbb{1}_{\{t>\tau \text{ and IO}\}}\right], \tag{3.9}$$

where τ is the date of occurrence of the single outlier, ω denotes the magnitude of the unique outlier and $\mathbb{1}_{\{t>\tau \text{ and IO}\}}$ the indicator function defined on a set of dates t, with $t \in [\tau, \ldots, T]$ and with e_τ being an IO.

The conditional variance $\hat{h}_t$ estimated from the noisy observed time-series e_t is as such:

$$\hat{h}_t = \alpha_0 + \beta_1 \hat{h}_{t-1} + \alpha_1 e_{t-1}^2. \tag{3.10}$$

If an outlier occurs at time τ, then from equation (3.8), for any $t \geq \tau$, we get the observed residual, $\hat{\nu}_t = e_t^2 - \hat{h}_t$, which is given such as (with previous notations):

$$\hat{\nu}_t = \nu_t + \omega \left[\mathbb{1}_{\{t=\tau\}} + \alpha_1 \left(\beta_1 - \alpha_1\right)^{t-\tau-1} \mathbb{1}_{\{t>\tau \text{ and IO}\}}\right]. \tag{3.11}$$

The expression (3.11) can be interpreted as a regression model for the unanticipated

volatility $\hat{\nu}_t$ (see Charles and Darné, 2005):

$$\hat{\mathbf{V}} = \omega \mathbf{X} + \mathbf{V}, \tag{3.12}$$

where $\hat{\mathbf{V}}$, $\mathbf{V}$ and $\mathbf{X}$ are some (T-dimensional) vectors, with T the length of the time-series, defined $\forall t \in [1, \ldots, T]$ as such:

$$\begin{cases} \hat{\mathbf{V}}_{[t]} = \hat{\nu}_t \text{ and } \mathbf{V}_{[t]} = \nu_t \\ \mathbf{X}_{[t]} = 0 + \mathbb{1}_{\{t=\tau\}} + \alpha_1 \left(\beta_1 - \alpha_1\right)^{t-\tau-1} \mathbb{1}_{\{t>\tau \text{ and IO}\}}. \end{cases}$$

The detection of the outlier can now start by computing the following two statistics:

$$\begin{cases} \hat{TS}_{AO}(\tau) & = \hat{\omega}\hat{\sigma}_{\hat{\nu}}^{-1} \left(\sum_{t=\tau}^{T} \mathbf{X}_{[t]}^2 \right)^{-1/2} \\ & = \left(\sum_{t=\tau}^{T} \mathbf{X}_{[t]} \hat{\nu}_t \right) \hat{\sigma}_{\hat{\nu}}^{-1} \left(\sum_{t=\tau}^{T} \mathbf{X}_{[t]}^2 \right)^{-1/2} \\ \hat{TS}_{IO}(\tau) & = \hat{\omega}\hat{\sigma}_{\hat{\nu}}^{-1} = \hat{\nu}_\tau \hat{\sigma}_{\hat{\nu}}^{-1}, \end{cases} \tag{3.13}$$

where $\hat{\sigma}_\nu$ denotes the estimated standard deviation of the residual.

Finally we evaluate $\hat{TS}_{AO}(\tau)$ and $\hat{TS}_{IO}(\tau)$ for all possible dates τ, with $\tau = [1, \ldots, T]$, and compute the following test-statistic:

$$\hat{TS}_{max} = \underset{\tau \in [1,\ldots,T]}{\text{Argmax}} \left\{ \hat{TS}_{AO}(\tau), \hat{TS}_{IO}(\tau) \right\}. \tag{3.14}$$

If the value of the statistic exceeds the predetermined critical value C, an outlier is detected at the observation for which $\hat{TS}_{max}$ is obtained. Based on extensive simulations, Chen and Liu (1993) propose a critical value of $C = 4$. The simulations we performed following the same process (but with another time-series) lead, in our case, to a critical value of $C = 10$ (as in Charles and Darné, 2005).

3.2.3 From Detection to Correction of Outliers

At this stage of the process, the potential outliers have now been detected. Various methods could then be applied to correct, remove and/or substitute these values. The first approach simply consists of removing the aberrant realizations from the original sample. To our knowledge, this is one of the oldest ways to deal with outliers (*Cf.* Bernoulli, 1778), and is still common nowadays (*Cf.* Chen and Liu, 1993). A second way to cope with these aberrant realizations is to consider them as missing values. A great number of methods to rebuild missing values exist in the literature. When dealing with asset returns, the Expectation-Maximisation algorithm is one of the main references (*Cf.* Dempster *et al.*, 1977). However, such a reconstruction failed to respect the higher-order (conditional) moment structure of the time-series under studies since they only focus on the first two moments. To overcome this problem, some methods have been recently proposed. For instance, Maillet and Merlin (2005), Merlin *et al.* (2009), Sorjamaa *et al.* (2009) propose various non-linear classification algorithms that lead to efficient completion of missing values. A third natural approach in our context consists of the use of a time-series structure and its implicit distributional assumption. We previously considered two kinds of structure: the traditional GARCH(1,1) model and the modified ANN-GARCH(1,1) that we will use in the following for evaluating the impact of outliers on portfolio selection.

In both approaches, the correction of outliers consists of the following steps. First, from the estimations of the GARCH(1,1) model and the observed series e_t, we obtain estimations of the conditional volatility series, denoted hereafter $\hat{h}_t$, of the unanticipated conditional volatility, such as $\hat{\nu}_t = e_t^2 - \hat{h}_t$, and of the outlier intensity, simply defined by $\hat{\omega} = \hat{\nu}_\tau$.

Thus, we replace e_t by e_τ^*, the GARCH-corrected return series, defined as (with previous notations):

$$e_t^* = sign(e_t) \left\{ e_t^2 - \hat{\omega} \left[\mathbb{1}_{\{t=\tau\}} + \alpha_1 \left(\beta_1 - \alpha_1 \right)^{t-\tau-1} \mathbb{1}_{\{t>\tau \text{ and IO}\}} \right] \right\}^{1/2}. \tag{3.15}$$

We finally return to the first step of the correction for the series e_t^*, and repeat all steps of the procedure until no $\hat{TS}_{max}$ test-statistic exceeds the critical value C.

If we now consider the ANN-GARCH time-series structure, the previous methodology can also be modified for an ANN-GARCH(1,1) volatility (hereafter denoted $\breve{h}_t$) model as (with previous notations):

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ \breve{h}_t = \alpha_0 + \beta_1 \breve{h}_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \mathfrak{M}\left(\breve{h}_{t-1}, \varepsilon_{t-1}\right). \end{cases} \tag{3.16}$$

In this case, the variable $\breve{\nu}_t$ – the unanticipated ANN-GARCH(1,1) volatility, can be expressed, like in the equation 3.8, such as (with previous notations):

$$\breve{\nu}_t = -\alpha_0 \left(1 - \beta_1\right)^{-1} + \sum_{k=0}^{t-1} \left\{ \pi_k \left[\varepsilon_{t-k}^2 + \mathfrak{M}\left(\breve{h}_{t-1-k}, \varepsilon_{t-1-k}\right) \right] \right\}. \tag{3.17}$$

In such a model, the extra non-linear component explained by the Artificial Neural Network is simply removed from the GARCH residual, the outlier intensity here being: $\breve{\omega} = \breve{\nu}_\tau$. The ANN-GARCH(1,1)-outlier corrected series $\breve{e}_t^*$ is then accordingly defined as (with previous notations):

$$\begin{aligned} \breve{e}_t^* = sign(e_t) \quad & \left\{ e_t^2 - \breve{\omega} \left\{ \mathbb{1}_{\{t=\tau\}} + \left[\alpha_1 \left(\beta_1 - \alpha_1\right)^{t-\tau-1} \right.\right.\right. \\ & \left.\left.\left. + \mathfrak{M}\left(\breve{h}_{t-1-k}, \varepsilon_{t-1-k}\right) \right] \mathbb{1}_{\{t>\tau \text{ and IO}\}} \right\} \right\}^{1/2}. \end{aligned} \tag{3.18}$$

As for the GARCH(1,1)-based procedure, we have then to return to the first step of the correction for the series e_t^* with the second detected outlier, and repeat all steps until no $\hat{TS}_{max}$ test-statistic exceeds the critical value C.

We have seen various approaches to detect and correct potential outliers. Before tackling the problem of evaluating the impact of outliers on Efficient Portfolio selection, we recall in the next section the higher-order moment asset allocation model setting, in the traditional as well as robust frameworks.

3.3 Portfolio Selection with Higher-order Moments

The mean-variance decision *criterion* proposed by Markowitz (1952 and 1991) is somehow inadequate for most of the risky asset allocation problems. Not only are the asset return distributions asymmetric and leptokurtic, but investors tend also to display preferences for positively skewed and light-tailed asset return distributions (see, for instance, Beedles and Simkowitz, 1978; Dittmar, 2002; Chang *et al.*, 2003; Semenov, 2004; Barberis and Huang, 2007; Agren, 2006). Different multi-moment approaches have been proposed in the financial literature to incorporate higher-order moment preferences into the asset allocation problems. We choose in the following to adopt the shortage function approach as in Briec *et al.* (2004) and Jurczenko *et al.* (2006 and 2008). The introduction of the shortage function enables us to optimize multiple objectives and consider alternatively classical as well as robust (to outliers) moments. We first recall the formulation of the classic comoments (denoted C-moments) and then present the robust statistic chosen: the linear moments (denoted L-moments).

3.3.1 Higher-order C-comoments of Portfolio Returns

In the following, we consider $\mathbf{R}$, a $(T \times N)$ matrix return database and a portfolio p invested in those N assets with weight w_{pi}, for $i = [1, \ldots, N]$. The mean, variance, skewness and *kurtosis* of portfolio p returns are given by:

$$\begin{cases} E\left(R_p\right) = \sum_{i=1}^{N} w_{pi}\, E\left(R_i\right) \\ \sigma^2\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^2\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{pi} w_{pj}\, \sigma_{ij} \\ m^3\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^3\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} w_{pi} w_{pj} w_{pk} m_{ijk} \\ \kappa^4\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^4\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} w_{pi} w_{pj} w_{pk} w_{pl} \kappa_{ijkl}, \end{cases} \tag{3.19}$$

with:

$$\begin{cases} \sigma_{ij} = E\left\{\left[R_i - E\left(R_i\right)\right]\left[R_j - E\left(R_j\right)\right]\right\} \\ m_{ijk} = E\left\{\left[R_i - E\left(R_i\right)\right]\left[R_j - E\left(R_j\right)\right]\left[R_k - E\left(R_k\right)\right]\right\} \\ \kappa_{ijkl} = E\left\{\left[R_i - E\left(R_i\right)\right]\left[R_j - E\left(R_j\right)\right]\left[R_k - E\left(R_k\right)\right]\left[R_l - E\left(R_l\right)\right]\right\}, \end{cases}$$

where w_{pi}, R_i, σ_{ij}, m_{ijk} and κ_{ijkl} represent, respectively, the weight of the asset i in portfolio p, the return on the asset i, the covariance between the returns on assets i and j, the coskewness between the returns on assets i, j and k, and the *cokurtosis* between the returns on assets i, j, k and l.

Based on the first writing proposed by Athayde and Florès (2002), Jurczenko *et al.* (2008) give a new vectorial notation allowing us to rewrite every higher co-moment matrix in a compact form with a recurrent relation. Thus, any (scalar) C-moment of order n, denoted $m^n\left(R_p\right)$, as well as any related global (higher-order) comoment $(N \times N^{n-1})$ matrix $\mathbf{M}^n$, with elements $\mathbf{M}^n_{(i,j)=(N\times N^{n-1})}$, with $j = \sum_{q=1}^{n-1}\left(a_{[q]}N^{n-1-q}\right)$, being such as, with $n \in [2, ..., 4]$ (and with previous notations):

$$\begin{cases} \underset{(1\times 1)}{m^n\left(R_p\right)} = & \mathbf{w}_p^{'} \times \mathbf{M}^n \times H^{[\bigcirc(n-2)]}\left(\mathbf{w}_p\right) \\ \underset{(1\times 1)}{m_{a_{[1]}\ldots a_{[n]}}} = & \mathbf{M}^n_{\left[a_{[1]},\sum_{q=1}^{n-1}\left(a_{[q+1]}-1\right)\times N^{n-1-q}\right]} = T^{-1}\mathbf{1}_T^{'}\left[\bigodot_{q=1}^{n}\check{\mathbf{R}}\left(a_{[q]}\right)\right], \end{cases} \tag{3.20}$$

with:

$$\begin{cases} \underset{(N\times 1)}{H^{(\bigcirc n)}\left(\mathbf{w}_p\right)} & = & \begin{cases} H^{[\bigcirc(n-1)]}\left(\mathbf{w}_p\right) \otimes \mathbf{w}_p' \text{ for } n > 0 \\ \mathbf{w}_p \text{ for } n = 0, \end{cases} \\ \underset{(T\times 1)}{\bigodot_{q=1}^{n}\check{\mathbf{R}}\left(a_{[q]}\right)} & \equiv & \underbrace{\check{\mathbf{R}}\left(a_{[1]}\right) \odot \ldots \odot \check{\mathbf{R}}\left(a_{[n]}\right)}_{n \text{ terms}}, \end{cases}$$

where the signs $\otimes$ and $\odot$ stand respectively for the symbol of the Kronecker and Hadamard product, and with $\check{\mathbf{R}}\left(q\right)$ the *q-th* column of $\check{\mathbf{R}}$, $\check{\mathbf{R}} = \mathbf{R} - \left(\mathbf{E} \times \mathbf{1}_T'\right)'$ being the $(T \times N)$ matrix of centered returns, $\mathbf{R}$ the $(T \times N)$ matrix of returns on the N assets, $\mathbf{1}_T$ the $(T \times 1)$ unit vector and the $a_{[q]}$ (with $q \in [1, ..., n] \subset [1, ..., N]$) being the ranks (column

number) of the assets in the matrix of excess returns $\check{\mathbf{R}}$ (taken in any order), that we want to compute.

3.3.2 Higher-order L-moments of Portfolio Returns

From the earlier work of Sillito (1951), Hosking (1989) provides a generalization of the L-moment computation. The L-moments are weighted linear sums of order statistics that are analogous to Conventional moments as simple descriptors of the shape of a general distribution. The use of L-moments covers (among others) the characterization of probability distributions and the summarization of observed data samples. Hosking and Wallis (1987) argue that L-moments provide better approximation of unknown distribution than C-moments. More specifically, when dealing with extreme realizations, L-moments lead to reliable estimations since they are more robust to outliers (*Cf.* Vogel and Fennessey, 1993). Indeed, they are only linearly influenced by large deviations and thus are more efficient than C-moments when characterizing distribution (*Cf.* Sankarasubramanian and Srinivasan, 1999; and Carrillo *et al.*, 2006). It is also to be noticed that L-moments always exist (whatever the order considered), whereas this is not true for traditional C-moments. The problem of working with non-defined quantities such as higher C-moments is then avoided. Finally, L-moments are also coherent shape measures of risk (*Cf.* Artzner *et al.*, 1999), since they are translation and scale invariant (*Cf.* Serfling and Xiao, 2007; and Gouriéroux and Jasiak, 2008).

In a general manner, L-moments can be expressed as a linear function of the Probability Weighted Moments, that is (see Hosking, 1989):

$$\lambda_k(X) = \sum_{r=1}^{k} p^*_{k-1,r-1}\, \beta_{r-1}(X), \tag{3.21}$$

with:

$$\beta_{r-1}(X) = r^{-1} E\left(X_{[r:r]}\right) = \int_0^1 Q(u)\; u^{r-1}\, du,$$

where $k \in \mathbb{N}^*$, $\beta_r(.)$ are the Probability Weighted Moments of order r, with $r = [2, ..., k]$,

and $p^*_{k,\,r}$ corresponds to the *r-th* coefficient of the shifted orthogonal Legendre polynomial of degree k denoted $P^*_k(.)$ and defined as $P^*_k(u) = P_k(2u-1)$ where $P_k(.)$ is the traditional Legendre polynomial of degree k.

Using the definition and orthogonal property of the shifted Legendre polynomials, the *k-th* population L-moment can also be represented as the covariance between the random variable X and its distribution function denoted $F(.)$, that is (see Serfling and Xiao, 2007):

$$\lambda_k(X) = \begin{cases} E(X) & \text{if } \; k = 1 \\ Cov\left\{X, P^*_{k-1}\left[F(X)\right]\right\} & \text{if } \; k \neq 1, \end{cases} \tag{3.22}$$

with:

$$E\left\{P^*_k\left[F(X)\right]\right\} = \int_0^1 P^*_k(u)\,du = 0,$$

where $P^*_0(u) = 1$, with $k \in IN^*$, $P^*_k(.)$ the shifted orthogonal Legendre polynomial of degree k defined as previously, $Cov(.,.)$ the covariance operator and $F(.)$ is the distribution of the random variable X.

Relying on the covariance representation of L-moments defined in Equation (3.22) and the bilinear property of the covariance operator, Serfling and Xiao (2007) introduced the L-(co)moments of multivariate variables. Jurczenko *et al.* (2008), relying on these last results, show that the first four L-moments of a portfolio returns could be expressed such as (with previous notations):

$$\begin{cases} \lambda_1\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\mathbf{E} \\ \lambda_2\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\boldsymbol{\Omega}^{(L)}_{\mathbf{w}_p} \\ \lambda_3\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\boldsymbol{\Sigma}^{(L)}_{\mathbf{w}_p} \\ \lambda_4\left(R_{\mathbf{w}_p}\right) = \mathbf{w}_p^{'}\,\boldsymbol{\Gamma}^{(L)}_{\mathbf{w}_p} - 3\,(2)^{-1}\,\mathbf{w}_p^{'}\,\boldsymbol{\Omega}^{(L)}_{\mathbf{w}_p}, \end{cases} \tag{3.23}$$

with:

$$\begin{cases} \underset{(1\times 1)}{E\left(R_{\mathbf{w}_p}\right)} & = \mathbf{w}_p^{'}\mathbf{E} \\ \underset{(N\times 1)}{\mathbf{\Omega}_{\mathbf{w}_p}^{(L)}} & = 2\,E\left\{\check{\mathbf{R}} \times \left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right) - E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}\right\} \\ \underset{(N\times 1)}{\mathbf{\Sigma}_{\mathbf{w}_p}^{(L)}} & = 6\,E\left\{\check{\mathbf{R}} \times \left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right) - E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}^2\right\} \\ \underset{(N\times 1)}{\mathbf{\Gamma}_{\mathbf{w}_p}^{(L)}} & = 20\,E\left\{\check{\mathbf{R}} \times \left\{F\left(\mathbf{R}_{\mathbf{w}_p}\right) - E\left[F\left(\mathbf{R}_{\mathbf{w}_p}\right)\right]\right\}^3\right\}, \end{cases}$$

where $F(.)$ is the distribution of the random variable $\mathbf{R}_{\mathbf{w}_p}$.

3.3.3 The Shortage Function and the Efficient Frontiers

The shortage function measures the distance between some points of the possible set of portfolio and the efficient frontier (*Cf.* Luenberger, 1995). The properties of the set of the portfolio return moments on which the shortage function is defined have already been discussed in the mean-variance plane by Briec *et al.* (2004) and in the higher-order moment space by Jurczenko *et al.* (2006) and Briec *et al.* (2007). Their definitions can be extended to obtain a portfolio efficiency indicator in the four L-moment framework (*Cf.* Jurczenko *et al.*, 2008). The shortage function associated with any portfolio $\mathbf{w}$ in the feasible set $\mathfrak{F} = \left\{\mathbf{w} \in \mathbb{R}^N : \mathbf{w}'\mathbf{1} = 1 \text{ and } \mathbf{w} \geq \mathbf{0}\right\}$, with reference to the direction vector $\mathbf{g} = (\, g_1\,;\, g_2\,;\, g_3;\, g_4)$, with $\mathbf{g} \in (\mathbb{R}^+ \times \mathbb{R}^- \times\ \mathbb{R}^+ \times \mathbb{R}^-) \setminus \{0\}$, in the four (robust) moment space, is the real-valued function $S_{\mathbf{g}}(.)$ defined as:

$$S_{\mathbf{g}}(\mathbf{w}) = \underset{\delta\in\mathbb{R}^+}{\text{Sup}}\left\{\delta : (\lambda_{\mathbf{w}} + \delta\mathbf{g}) \in \mathfrak{F}\right\}. \tag{3.24}$$

Let us consider a specific portfolio p, defined by its vector of weights denoted $\mathbf{w}_p$, composed of a set of N assets and whose performance needs to be evaluated in the four (L-)moment dimensions. The portfolio optimization problem corresponding to the investor willing to

improve the efficiency of its portfolio in the direction $d = (d_1, d_2, d_3, d_4)$ is:

$$\mathbf{w}^* = \underset{(\mathbf{w},\,\delta)\in(\mathfrak{F}\times\,\mathbb{R}^+)}{\text{Argmax}} \{\delta\} \quad s.t. \begin{cases} M_1\left(R_{\mathbf{w}_p}\right) + \delta\, d_1 \leq M_2\left(R_{\mathbf{w}}\right) \\ M_2\left(R_{\mathbf{w}_p}\right) - \delta\, d_2 \geq M_2\left(R_{\mathbf{w}}\right) \\ M_3\left(R_{\mathbf{w}_p}\right) + \delta\, d_3 \leq M_3\left(R_{\mathbf{w}}\right) \\ M_4\left(R_{\mathbf{w}_p}\right) - \delta\, d_4 \geq M_4\left(R_{\mathbf{w}}\right), \end{cases} \tag{3.25}$$

where M_i for $i = [1, \ldots, 4]$ correspond to the moments (robust or not) of order i.

From equation (3.20), the C-moments of the portfolio could be written as (with previous notations):

$$\begin{cases} E\left(R_p\right) & = \mathbf{w}_p^{'}\,\mathbf{E}\times 1 \\ \sigma^2\left(R_p\right) & = \mathbf{w}_p^{'} \times \mathbf{\Omega}_{\mathbf{w}_p} & = \mathbf{w}_p^{'} \times \left(\mathbf{\Omega} \times \mathbf{w}_p\right) \\ m^3\left(R_p\right) & = \mathbf{w}_p^{'} \times \mathbf{\Sigma}_{\mathbf{w}_p} & = \mathbf{w}_p^{'} \times \left\{\mathbf{\Sigma}\times \left[H^{(\bigcirc 1)}\left(\mathbf{w}_p\right)\right]\right\} \\ \kappa^4\left(R_p\right) & = \mathbf{w}_p^{'} \times \mathbf{\Gamma}_{\mathbf{w}_p} & = \mathbf{w}_p^{'} \times \left\{\mathbf{\Gamma}\times H^{(\bigcirc 2)}\left(\mathbf{w}_p\right)\right\}. \end{cases} \tag{3.26}$$

The previous optimization problem considering C-moments is then equivalent to (with previous notations):

$$\mathbf{w}^* = \underset{(\mathbf{w},\,\delta)\in(\mathfrak{F}\times\,\mathbb{R}^+)}{\text{Argmax}} \{\delta\} \quad s.t. \begin{cases} \mathbf{w}_p'\mathbf{E} + \delta\, d_1 \leq \mathbf{w}'\mathbf{E} \\ \mathbf{w}_p'\mathbf{\Omega}_{\mathbf{w}_p} + \delta\, d_1 \geq \mathbf{w}'\mathbf{\Omega}_{\mathbf{w}} \\ \mathbf{w}_p'\mathbf{\Sigma}_{\mathbf{w}_p} + \delta\, d_1 \leq \mathbf{w}'\mathbf{\Sigma}_{\mathbf{w}} \\ \mathbf{w}_p'\mathbf{\Gamma}_{\mathbf{w}_p} + \delta\, d_1 \geq \mathbf{w}'\mathbf{\Gamma}_{\mathbf{w}}. \end{cases} \tag{3.27}$$

From Equation 3.23, the previous optimization problem can easily be rewritten in a

robust L-moment framework as such (with previous notations):

$$\mathbf{w}^* = \underset{(\mathbf{w},\,\delta)\in(\mathfrak{F}\times\,\mathbb{R}^+)}{\text{Argmax}}\{\delta\}$$
$$s.t.\begin{cases}\mathbf{w}_p'\mathbf{E}+\delta\,d_1\leq\mathbf{w}'\mathbf{E}\\ \mathbf{w}_p'\boldsymbol{\Omega}^{(L)}_{\mathbf{w}_p}+\delta\,d_1\geq\mathbf{w}'\boldsymbol{\Omega}^{(L)}_{\mathbf{w}}\\ \mathbf{w}_p'\boldsymbol{\Sigma}^{(L)}_{\mathbf{w}_p}+\delta\,d_1\leq\mathbf{w}'\boldsymbol{\Sigma}^{(L)}_{\mathbf{w}}\\ \mathbf{w}_p'\boldsymbol{\Gamma}^{(L)}_{\mathbf{w}_p}-3/2\mathbf{w}_p'\boldsymbol{\Omega}^{(L)}_{\mathbf{w}_p}+\delta\,d_1\geq\mathbf{w}'\boldsymbol{\Gamma}^{(L)}_{\mathbf{w}}-3/2\mathbf{w}'\boldsymbol{\Omega}^{(L)}_{\mathbf{w}}.\end{cases} \tag{3.28}$$

Since the various higher-order moments optimization program (classical and robust) are now defined, we can now evaluate the impact of outliers on such asset allocation models.

3.3.4 Distributional Timing and Higher-order Moments

When dealing with investors' higher-order moment preferences and asset allocation models, the optimization of a utility function (approximated to the fourth order by a Taylor expansion) leads to *optima* that are close to those obtained with the Mean-Variance *criterion* (*Cf.* Levy and Markowitz, 1979; Kroll *et al.*, 1984). These last results suggest that higher-order moments have no significant impact on the investor's portfolio choice. However, Jondeau and Rockinger (2006) as well as Jurkzenko *et al.* (2008) report that for highly asymmetric and fat-tailed distributions, the mean-variance *criterion* may still fail to approximate investor preferences. Jondeau and Rockinger (2009) also emphasize the importance of higher-order moments in a dynamic asset allocation framework. In a dynamic setting indeed, asset return distributions may vary, leading to partial-short term forecasts. These dynamic properties have been applied to the first two moments. The concept of market and volatility timing stands for the idea that when doing portfolio asset allocation, investors use return and volatility forecasts (*Cf.* Kandel and Stambaught, 1996; Barberis, 2000, among others). Several empirical studies have shown that the volatility timing is of significant economic value for investing (*Cf.* for instance Fleming *et al.*, 2001). Simulations of strategies based on volatility forecasts highlight the fact that taking into account volatility variability is valuable for investors.

Jondeau and Rockinger (2009) generalize this timing concept to the first four moments and call it the "distributional timing". To reveal the impact of higher-order moment evolution on asset allocation, dynamics of moments are captured with a conditional model for higher-order moments. Extending the GARCH model for conditional skewness and *kurtosis* of Harvey and Siddique (1999), Jondeau and Rockinger (2006) first consider a generalized t-Student Distribution that models asymmetry and *kurtosis* within a more flexible parameter scheme. Jondeau and Rockinger (2008) secondly evaluate an extension of the Dynamic Conditional Correlation model of Engle (2002) with a Skew t-Student multivariate distribution (*Cf.* Sahu *et al.*, 2003) for the innovation, modelling ε_t as such (with previous notations):

$$\begin{cases} \varepsilon_t = \mathbf{\Sigma}_t^{1/2} \eta_t \\ \eta_t \sim g\left(\eta_t \,|\theta_t\right) \end{cases} \tag{3.29}$$

where $\mathbf{\Sigma}_t$ stands for the conditional covariance matrix and $g\left(.\,|\theta_t\right)$ is a Skew t-Student multivariate distribution of η_t with time-varying shape parameters $\theta_t = [\theta_{1,\,t}\, \theta_{2,\,t}]$.

Their model leads to a realistic higher-order moment characterization with both a dynamically adjustable asymmetry and a variable tail thickness. Coefficients driving the conditional skewness and *kurtosis* are modelled as such (with previous notations):

$$\begin{cases} \theta_{1,\,t} = & c_0 + c_1^- \left|\eta_{t-1}\right| \mathbb{1}_{\{\eta_{t-1}<0\}} + c_1^+ \left|\eta_{t-1}\right| \mathbb{1}_{\{\eta_{t-1}\geq 0\}} + c_2 \theta_{1,\,t-1} \\ \theta_{2,\,t} = & d_0 + d_1^- \left|\eta_{t-1}\right| \mathbb{1}_{\{\eta_{t-1}<0\}} + d_1^+ \left|\eta_{t-1}\right| \mathbb{1}_{\{\eta_{t-1}\geq 0\}} + d_2 \theta_{2,\,t-1} \end{cases} \tag{3.30}$$

where $\theta_{1,\,t}$ and $\theta_{2,\,t}$, for $t = [1, \ldots, T]$, govern respectively the asymmetry and the fat-tailedness feature of the distribution, $|.|$ stands for the absolute value operator, $\mathbb{1}_{\{.\}}$ is the indicator function, and c_0, c_1^-, c_1^+, c_2, d_0, d_1^-, d_1^+ and d_2 are real coefficients.

From their moment forecasts, they use a utility-based asset allocation model to evaluate the usefulness of the "distributional timing". They show that distributional timing is important for the asset allocation and that, in a dynamic setting, the higher-order comoment computations have some economic value.

However, some issues may limit the performance of such a model. First, even if the t-Student hypothesis leads to a time-series structure that respects the third and fourth moments, the global shape of asset return distributions may differ significantly from the chosen t-Student (specifically in the center of the distribution where the most important part of the returns belongs). Secondly, the extreme comoments might not be well-modelled since the parametrization of the t-Student does not take into account any dependence of extremes. Thirdly, this last approach suffers from the curse of dimensionality; for large scale optimization problems, it is just computationally infeasible.

León *et al.* (2005) propose an alternative approach to model conditional higher-order moments, which is more practical since the model specification and the estimation step are simplified. Their model is thus estimated assuming a Gram-Charlier series expansion of the Normal density function for the innovation term. Such an expansion enables us to take into account the departure from Gaussiantity of the innovation term and leads to an easier estimation than the one of a non-central t-Student distribution. The estimation is made within a two-step procedure. Assuming a traditional Gaussian GARCH$(1,1)$ model, parameters are first estimated using the classical Maximum Likelihood method. If then the necessary conditions on parameters are verified (finite variance and existence of first four moments, Cf. He and Teräsvirta, 1999), a second Maximum Likelihood optimization is performed on a more complete model (using the first step parameter estimates as a starting point), to finally to obtain parameters governing the conditional skewness and *kurtosis*. Similarly, we propose hereafter to use an Artificial Neural Network to extend the conditional higher-order moment modelling, and adopt in a dynamic framework an Artificial Neural Network - Generalized AutoRegressive Moment model (denoted ANN-GARM in the following). The extended ANN-GARM(p,q) model takes the general form (with previous notations):

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t} \\ \mathbb{M}_t = \boldsymbol{\alpha}^0 + \sum_{i=1}^{p} \boldsymbol{\beta}^i \mathbb{M}_{t-i} + \sum_{j=1}^{q} \boldsymbol{\alpha}^j \varepsilon_{t-j} + \mathring{\mathfrak{M}}\left(\mathbf{I}_{t-1}\right) \end{cases} \tag{3.31}$$

where, for $t = [1, \ldots, T]$, $\mathbb{M}_t = [\sigma^2(\varepsilon)\ m^3(\varepsilon)\ \kappa^4(\varepsilon)]'$, $\varepsilon_t = [\varepsilon_t^2\ \varepsilon_t^3\ \varepsilon_t^4]'$, $\mathbf{I}_{t-1} = [\mathbb{M}_{t-1}, \ldots, \mathbb{M}_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}]$ and, for $i = [1, \ldots, q]$, $\boldsymbol{\alpha}^i = [\alpha_1^i\ \alpha_2^i\ \alpha_3^i]'$, for $i = [1, \ldots, p]$, $\boldsymbol{\beta}^i = [\beta_1^i\ \beta_2^i\ \beta_3^i]'$, and where $\mathring{\mathfrak{M}}(.)$ is an MLP with three output neurons aiming to estimate the unanticipated volatility, skewness and *kurtosis*.

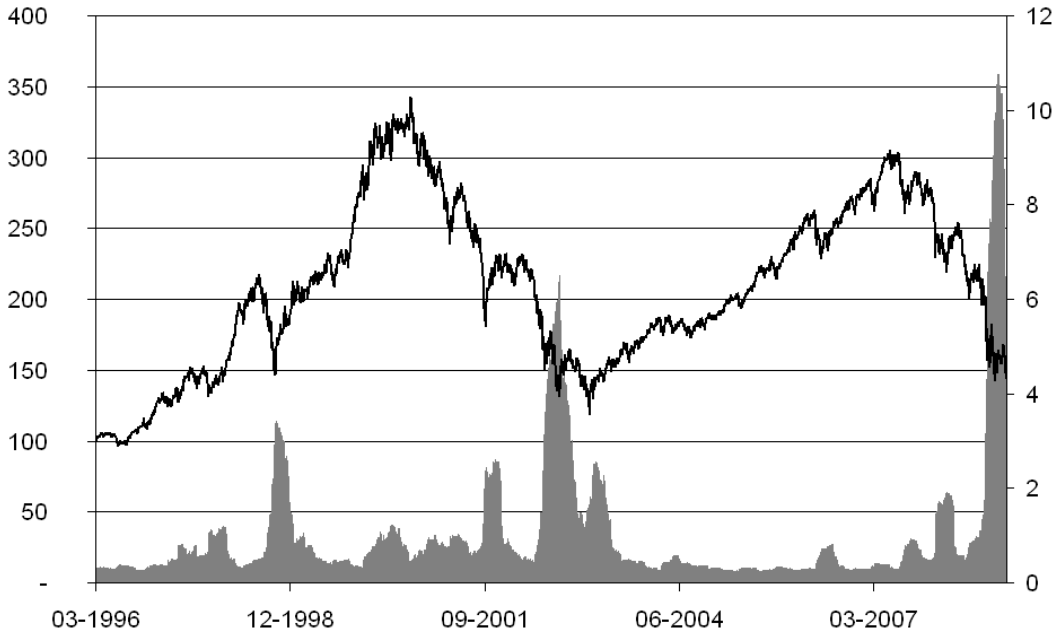
Based on a GARM$(1, 1)$ specification of the conditional higher-order moment dynamics, we should be able to use moment forecasts to improve out-of-sample properties of optimized portfolios, thanks to a better risk assessment (*Cf.* Fleming *et al.*, 2001; Dubauskas and Teresiené, 2005; and Jondeau and Rockinger, 2009).

3.4 Data and Empirical Results

In the following empirical application, we explore a dataset of quotes of stocks included in the CAC40 Index. The time-series of the market Index constituents are provided by Datastream on a daily basis. The considered period is from January 1996 to January 2009. The dataset consists of daily Euro (converted from French Franc before 2000) denominated returns of a sample of 29 stocks included in the CAC40 for the global period. Since we needed a complete dataset, we thus simply removed from the sample the 11 stocks that were included in the Index after January 1996.

Figure 3.2 represents the CAC40 Index evolution. The considered period is well adapted to the studies of outliers and extreme events. The period indeed includes more than a complete economic cycle and is characterized by a succession of bear and bull markets. More specifically, we can find in the sample period events such as the Asian (1997) and Russian (1998) crises, the Internet bubble boom and crash, the 9/11/2001 event, the market correction of May 2006 and the recent Subprime crisis. In order to differentiate systematic risk from specific risks (company or sector), we also consider in the following the Index of Market Shocks (IMS for short, *Cf.* Maillet and Michel, 2003 and 2005). This measure relies on an analogy with geophysics: the IMS is equivalent to the Richter scale used for earthquakes. More precisely, as a market is the place where economic agents with

Figure 3.2. French Equity Market Evolution



Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. This figure represents French Equity market evolution (black thin line corresponding to the left y-axis) as well as the Indicator of Market Shocks (grey area corresponding to the right y-axis).

different investment horizons exchange, the IMS definition is a weighted aggregation of filtered volatility measures, corresponding to different horizons of the interacting investors. The comparisons of CAC40 Index-based IMS and a specific stock-based IMS provide a way to differentiate specific from systematic shocks. Figure 3.2 shows the evolution of the CAC40 Index-based IMS.

We also performed some additional tests on stock return Gaussianity. Tables 3.13 and 3.14 in Appendix 3.6.2 presents the results of traditional (uncorrected) tests, at a 5% significance threshold. Unsurprisingly, the Gaussian hypothesis is rejected and a GARCH(1,1) modelling is justified for all series.

3.4.1 About the Detection of Market Data Outliers on the French Market

We first start by applying the GARCH and ANN-GARCH outlier detection and correction methods on the original set of (raw) returns. We find 174 outliers when using an

ANN-GARCH and 178 outliers with a GARCH approach. The first 174 GARCH detected outliers correspond exactly to those detected with the ANN-GARCH. Due to a better volatility modelling with the ANN-GARCH, the residual and the associated test-statistic are lower[3]. Thus, even if the ANN-GARCH leads to an 11% average improvement of Root Mean Squared Errors of the volatility modelling, we cannot conclude that ANN-GARCH modelling significantly improves the outlier detection process.

Figure 3.3 illustrates the most significant outliers detected with both methods. Outlier intensities are plotted on the left axis whereas the differences between the two estimated outlier intensities are plotted on the left axis. Only minor differences are observed between the two outlier detection procedures. The addition of the artificial neural network improves the volatility modelling. As a consequence, the ANN-GARCH process estimates outlier with lower intensity. Significant differences are observed for return time-series with high *kurtosis*.

We observe that the correction process has a strong impact on higher moments and clearly reduce the departure from Gaussianity of stock returns (see Table 3.1 and Table 3.12 for complete results).

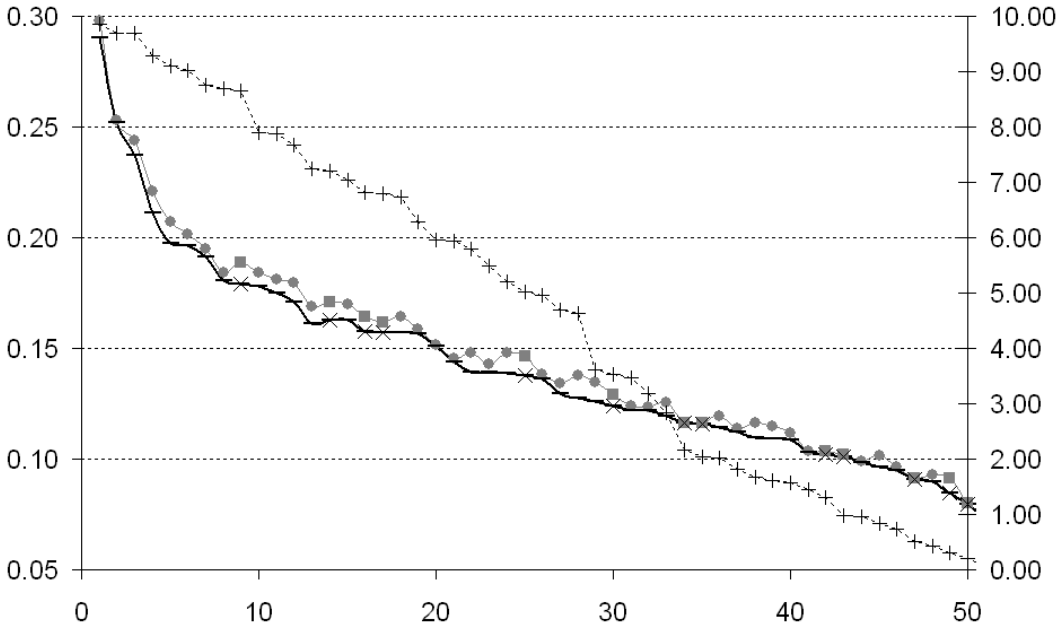
Table 3.1. Descriptive Statistics

Series	Type	Moment 1*	Moment 2*	Moment 3	Moment 4
ALCATEL-	Unadjusted	-1.71	54.22	.17	16.97
LUCENT	Normal GARCH Adjusted	-3.37	51.43	-.33	7.17
	ANN-GARCH Adjusted	-3.41	52.64	-.12	6.66
PERNOD-	Unadjusted	14.28	32.04	.22	7.98
RICARD	Normal GARCH Adjusted	11.65	30.45	.27	6.64
	ANN-GARCH Adjusted	11.02	30.22	.27	6.63
VIVENDI	Unadjusted	6.40	39.46	-.48	20.35
	Normal GARCH Adjusted	3.79	30.45	.05	6.37
	ANN-GARCH Adjusted	3.76	35.25	.04	5.97

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. This Table presents the first four moments of several stocks when the data are raw (Unadjusted), corrected with the method proposed by Charles and Darné (2005, Normal GARCH Adjusted) or corrected with the MLP enhanced method we propose (ANN-GARCH Adjusted). *Moments are annualized and expressed in %.

[3]The four outliers neglected by the ANN-GARCH detection have a test-statistic just below the critical value of the chosen threshold equal to 10.

Figure 3.3. Comparison of ANN-GARCH and GARCH detected Outliers



Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. This figure represents the intensity of the most significant outliers detected with GARCH (in the grey line) and ANN-GARCH models (in the black line) on the left axis. Big dots, squares, ticks and crosses represent respectively GARCH detected AO, GARCH detected IO, ANN-GARCH detected AO and ANN-GARCH detected IO. On the right axis are represented the differences of intensity between ANN-GARCH and GARCH detected Outliers multiplied by 1,000 and ploted in the dashed grey line.

Table 3.2 provides descriptive statistics of the detected outliers. Every time-series from the dataset is impacted by outliers. We find that Additive Outliers appear twice as often as Innovative Outliers (see Table 3.2). The breakdown between positive and negative outliers seems well balanced. We also observed that Innovative Outliers (Additive Outliers) are more related to negative (positive) returns, which illustrate the well-known asymmetric behavior of volatility (*Cf.* Glosten *et al.*, 1993). We find an average number of 6 outliers for each time series with high dispersion depending on the considered asset (from 1 to 18).

Table 3.3 presents some of the most significant detected outliers. From this restricted set, only one outlier is found to be innovative (concerning the stock Vivendi on June the 24^{th} of 2002). Most of these associated returns are negative, which seems in accordance with the asymmetric effect of "news" proposed by Black (1976). For each observed outlier,

Table 3.2. Descriptive Statistics of the Outliers

		Outlier Intensities			T-statistics		
		Mean	Minimum	Maximum	Mean	Minimum	Maximum
GARCH	AO	8.47	3.42	29.79	17.14	12.01	45.52
	IO	12.79	3.89	15.74	15.99	12.07	31.62
ANN-	AO	8.23	3.41	29.03	16.58	12.00	45.37
GARCH	IO	12.21	5.56	17.92	15.76	12.07	31.68

		Outliers		Number of Series	Number of Outliers in Series		
		Positive	Negative	subject to Outliers	Mean	Minimum	Maximum
GARCH	AO	76	46	29	4.13	1	18
	IO	17	39	24	1.86	0	5
ANN-	AO	74	45	29	4.06	1	17
GARCH	IO	17	38	22	1.81	0	5

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors.

we try to evaluate the cause and explanation of the extreme events. A first qualitative assessment is made. Using past Bloomberg^{TM} market reports, we try to find any relevant information that could explain the outlier. Then, this analysis is completed with a quantitative one. We compare the values of the Indicator of Market Shocks obtained with the CAC40 Index with the one computed from the stock specific return time-series. If only the stock-based specific IMS is high, then one can speculate about the realization of a specific risk that day. From Table 3.3, very few (idiosyncratic) outliers appear during the major financial crises (signaled by a value greater than 3.48, corresponding to a 95% threshold of the IMS distribution; *Cf.* Maillet and Michel, 2005).

Table 3.3. Largest Outliers on the Database

Date	Return	Company	t-stat	Type	Explanation	IMS
1. 09/17/1998	-38.40%	ALCATEL-LUCENT	45.52	AO	Crash of telecom sector	7.92 [1.30]
2. 01/07/2005	-16.77%	UNIBAIL-RODAMCO	42.30	AO	Exceptional coupon	6.18 [.29]
3. 07/26/2002	-22.05%	SAINT GOBAIN	37.18	AO	Quarter results	5.13 [3.55]
4. 07/03/2002	-21.91%	VIVENDI	35.63	AO	CEO resigned	8.70 [1.47]
5. 10/30/2002	40.54%	ALCATEL-LUCENT	35.13	AO	Quarter results	5.14 [4.73]
6. 09/28/2001	-20.40%	SCHNEIDER ELECTRIC	34.70	AO	Failure of the Legrand merger	6.69 [2.33]
7. 01/21/2005	25.80%	VALLOUREC	32.21	AO	Takeover V & M Tubes	1.68 [.26]
8. 06/24/2002	-23.31%	VIVENDI	31.68	IO	Key board member resigned	3.64 [1.05]
9. 10/17/1996	23.66%	LAGARDERE GROUPE	29.95	AO	Takeover Thomson	1.65 [.27]
10. 06/26/2001	-22.93%	CAP GEMINI	27.72	AO	Profit Warning	5.53 [1.12]
11. 07/02/2002	-25.52%	VIVENDI	27.42	IO	Rumour that CEO resigned	5.08 [1.12]
12. 10/17/2002	26.38%	CAP GEMINI	23.75	AO	Profit Warning	6.39 [5.25]
13. 12/17/2008	-17.24%	BNP PARIBAS	23.30	AO	Profit Warning	9.97 [10.44]
14. 11/21/2002	11.91%	UNIBAIL-RODAMCO	22.92	AO	Change in fiscal system	1.16 [3.77]
15. 10/13/2008	18.12%	VINCI (EX SGE)	22.02	IO	Market turmoil	.96 [4.93]

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors.

3.4.2 About Financial Features of some Notorious Efficient Portfolios

Regarding the portfolio allocation problem, we now want to evaluate the impact of outliers on efficient portfolios. Considering, for instance, extreme higher-order moment portfolios, such as the Global Minimum *Kurtosis* portfolio, may lead to the selection of some unrealistic portfolios. Indeed, Jondeau and Rockinger (2003, among others) observe that higher-order moments have only a marginal impact on portfolios that maximize utility function.

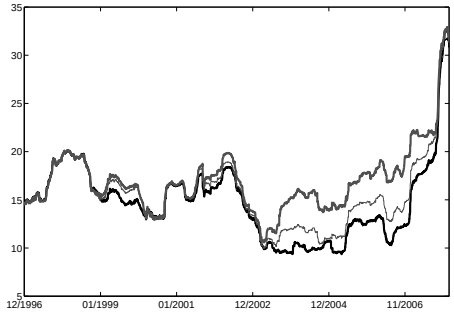
However, as pointed out by Jurczenko *et al.* (2008), a potential application for the investors of higher (L-)moments may lie in controlling the features of optimal portfolios, since some efficient portfolios, similar in terms of the first two moments, exhibit large differences in higher-order L-moments. Thus, we choose to restrict the set of optimal portfolios to those with a volatility that is not significantly different from that of the Global Minimum Volatility Portfolio ($GMVP$) at a 5% threshold using the Log-normal hypothesis of Alizadeh *et al.* (2002). We also select the Local Maximum Mean ($LMMP$), the Local Maximum (L-)moment$_3$ ($LMSP$ and LML_3P) and the Local Minimum (L-)moment$_4$ Portfolios ($LMKP$ and LML_4P)[4]. One way to remain in this restricted portfolio set is to integrate this constraint in the optimization process. For each optimization, we first compute the composition of the Minimum Volatility Portfolio and the distribution of its volatility. The associated confidence interval is then added to the second moment constraint.

Using a one-year rolling window, we perform a dynamic asset allocation with a monthly reallocation. Note that the outlier corrected database is only used for optimal weight determination (all the following statistics are based on the true database). Figure 3.4 shows the evolution of (L-)Moments of the notorious portfolios. The details of moment peculiarities are presented hereafter but we can already see that L-moments are clearly more

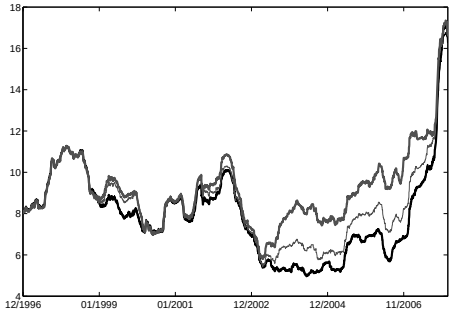
[4]In the following, for the sake of simplicity and to differentiate portfolios obtained with unadjusted returns from those corresponding to corrected-from-outliers series, we add an extra $\breve{}$ onto the name of the portfolios when they are obtained with corrected returns. For instance the Local Minimum L-moment$_4$ Portfolios are denoted LML_4P and $LM\breve{L}_4P$.

stable than the traditional C-moments. The influence of outliers is evaluated according to the *ex post* observed moments of these particular portfolios. We also consider additional descriptive statistics of portfolios such as the traditional Sharpe ratio, the Concentration Coefficient (denoted CC, *Cf.* Gillman, 2004), and a measure of the allocation stability which is the Turn-over Ratio (denoted TR). Since the Sharpe ratio does not take into account the non normality of the distribution, we also consider utility based measures (*Cf.* Appendix 3.6.3). For each local optimization, we select the portfolio optimized with raw returns and C-moments as benchmark and compute the cost or management fee an investor is willing to pay to switch from the benchmark portfolio to robust dynamic strategy. These utility measures are computed in two frameworks: the traditional Expected Utility and the Cumulative Prospect Theories. Indeed, the traditional Expected Utility Theory hardly takes into account higher order moment preferences (*Cf.* Jurczenko *et al.*, 2008). The Prospect Theory provides an efficient alternative to account for numbers of anomalies (*Cf.* Ingersoll, 2008) such as the skewness preference of investors. A weighting function is applied to the cumulative probability distribution of the portfolio returns. The main effect of this transformation is to make the agent overweigh the tails of that distribution.

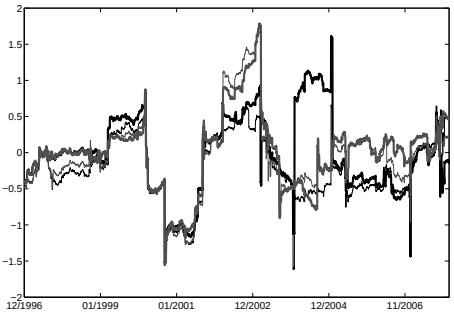
Figure 3.4. Time-evolutions of some Notorious Efficient Portfolios



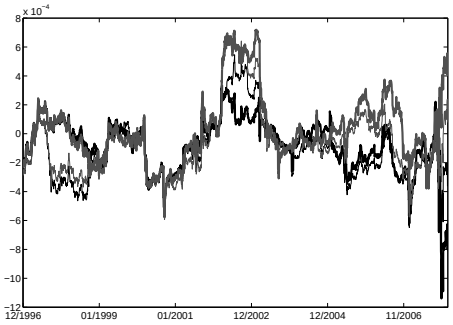
Volatility of $GMVP$, $G\breve{M}VP$, GML_2P and $G\breve{M}L_2P$



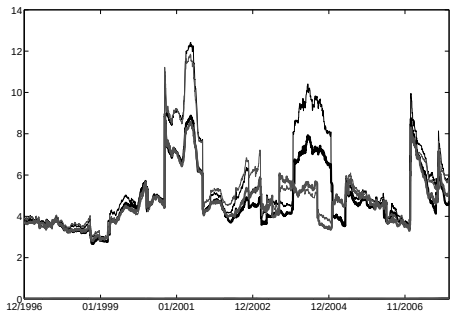
L-moment$_2$ of $GMVP$, $G\breve{M}VP$, GML_2P and $G\breve{M}L_2P$



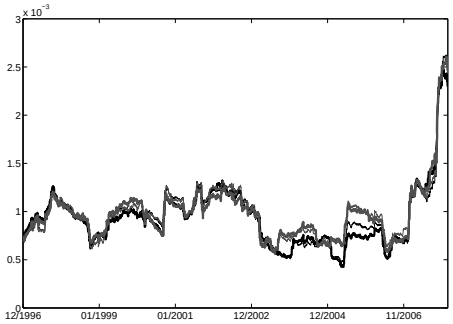
Skewness of $LMSP$, $LM\breve{S}P$ LML_3P and $LM\breve{L}_3P$



L-moment$_3$ of $LMSP$, $LM\breve{S}P$ LML_3P and $LM\breve{L}_3P$



Kurtosis of $LMKP$, $LM\breve{K}P$ LML_4P and $LM\breve{L}_4P$



L-moment$_3$ of $LMKP$, $LM\breve{K}P$ LML_4P and $LM\breve{L}_4P$

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. We use a one-year rolling window to evaluate the time evolution of the moments. In each subgraph, the black thin line corresponds to the portfolio obtained with unadjusted (raw) returns and a C-moment optimization, the black bold line corresponds to the portfolio obtained with unadjusted (raw) returns and an L-moment optimization, the grey thin line corresponds to the portfolio obtained with ANN-GARCH-adjusted returns and a C-moment optimization and the grey bold line corresponds to the portfolio obtained with ANN-GARCH-adjusted returns and an L-moment optimization. $LMMP$, $GMVP$, GML_2P, $LMSP$, LML_3P, $LMKP$ and LML_4P stands for the Local Maximum Mean, the Global Minimum Volatility, the Global Minimum L-moment$_2$, the Local Maximum Skewness, the Local Maximum L-moment$_3$, the Local Minimum *Kurtosis* and the Local Minimum L-moment$_4$ Portfolio. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

Table 3.4 presents the values of ratios for the Local Maximum Mean Portfolio. In this case, we do not display statistics corresponding to L-moment portfolio optimization since the C-moment$_1$ and L-moment$_1$ are strictly equivalent. Using adjusted data leads to an improvement of every moment except for volatility. The two portfolios are very close regarding their volatility (with a 35 basis point difference), but when correcting for the outliers, expected return significantly increases whilst skewness becomes positive and extreme risk, measured by the *kurtosis*, is reduced by a fourth. Regarding the performance measures (see Table 3.5), the portfolio obtained after data correction is preferred whatever the ratio considered. Adjusting the returns from outliers leads to a more concentrated allocation and the turn-over of the two portfolios are quite equivalent. These results indicate that the first moment is highly subject to bias and optimizations in which expected return matters could also be improved by adjusting the data from outliers.

Table 3.4. Local Maximum Mean Portfolios

	Moment 1	Moment 2	Moment 3	Moment 4
$LMMP$	9.66	18.97 [9.24]	-.04 [-.27]	20.66 [1.46]
$LM\breve{M}P$	12.43	19.32 [9.33]	.18 [-.14]	15.18 [1.49]

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. The first two moments are annualized and expressed in %. The corresponding values of L-moments are indicated in brackets. L-moments 3 and 4 are multiplied by 1,000. $LMMP$ corresponds to the Local Maximum Mean Portfolio. We add an extra $\breve{}$ onto the name of the portfolio when it is obtained with corrected returns.

Table 3.5. Performance Measures for Local Maximum Mean Portfolios

	SR	CC	TR	WP_2	WP_{10}	WP_2^*	WP_{10}^*
$LMMP$	.35	6.92	1.63	.00	.00	.00	.00
$LM\breve{M}P$	.48	5.39	1.65	2.53	2.44	21.85	18.75

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. SR stands for the Sharpe Ratio, CC is for the Concentration Coefficient and TR denotes the Turn-over Ratio. The Willing-to-Pay measures are denoted WP_γ and WP_γ^* when they are computed respectively in the traditional Expected Utility and in the Cumulative Prospect Theory frameworks, γ being the level of the risk aversion. The corresponding values of the Willing-to-Pay measures (compared to the $LMMP$) are annualized and expressed in percent. $LMMP$ corresponds to the Local Maximum Mean Portfolio. We add an extra $\breve{}$ onto the name of the portfolio when it is obtained with corrected returns.

When focusing on Global Minimum Risk Portfolios (see Table 3.6), whatever the robust approach chosen (L-moments or ANN-GARCH Adjusted), the volatility is deteriorated. This empirical observation is in accordance with previous empirical studies that conclude that the minimum volatility portfolio is robust to outliers (see for instance, Jorion, 1986; Michaud, 1988; Chopra and Ziemba, 1993; Jagannathan and Ma, 2003; and Ledoit and Wolf, 2003). Now regarding the use of the L-moments, in both cases, it increases the risk of the portfolios. A potential explanation is that L-moments are said to be robust to outliers because they are stable (or not too sensitive to new outcomes). So when dealing with dynamic asset allocation, L-moments may not be reactive enough to capture short term market evolutions. Regarding the others moments, the $GM\breve{L}_2P$ exhibits a higher expected return and skewness while the portfolios obtained from adjusted returns have improved third and fourth moments but lower first moments. Consequently, performance measures that mostly focus on the first two moments (the Sharpe ratio as well as the Willing-to-Pay ratio in the Expected Utility framework) are lower for portfolios obtained with adjusted returns. In the opposite, if we consider ratios that emphazise the higher-order moments, the three robust portfolios seem better than the $GMVP$. The use of an ANN-GARCH return correction leads to more concentrated portfolios, and an optimization with L-moments provides more stable allocations.

Table 3.6. Global Minimum Volatility Portfolios

	Moment 1	Moment 2	Moment 3	Moment 4
$GMVP$	13.87	16.57 [8.60]	-.22 [-.17]	9.30 [1.25]
GML_2P	14.52	16.71 [8.63]	-.11 [-.16]	9.80 [1.28]
$G\breve{M}VP$	13.21	17.28 [9.04]	.12 [-.14]	8.66 [1.26]
$GM\breve{L}_2P$	12.58	18.42 [9.68]	.01 [-.06]	7.89 [1.36]

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. The first two moments are annualized and expressed in %. The corresponding values of L-moments are indicated in brackets. L-moments 3 and 4 are multiplied by 1,000. $GMVP$ and GML_2P correspond to the Global Minimum Volatility and Global Minimum L-moment$_2$. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

We now move to the study of the influence of outliers on higher-order moments and the related optimized portfolios. Table 3.8 presents the results obtained through the four

Table 3.7. Performance Measures for Global Minimum Volatility Portfolios

	SR	CC	TR	WP_2	WP_{10}	WP_2^*	WP_{10}^*
$GMVP$	.65	7.07	1.08	.00	.00	.00	.00
GML_2P	.68	7.31	.97	.58	.48	7.81	6.56
$G\breve{M}VP$	.59	5.83	1.04	-.97	-1.76	4.68	3.12
$G\breve{M}L_2P$	.51	5.34	.87	-1.95	-4.39	14.06	9.37

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. SR stands for the Sharpe Ratio, CC is for the Concentration Coefficient and TR denotes the Turn-over Ratio. The Willing-to-Pay measures are denoted WP_γ and WP_γ^* when they are computed respectively in the traditional Expected Utility and in the Cumulative Prospect Theory frameworks, γ being the level of the risk aversion. The corresponding values of the Willing-to-Pay measures (compared to the $GMVP$) are annualized and expressed in percent. $GMVP$ and GML_2P correspond to the Global Minimum Volatility and Global Minimum L-moment$_2$. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

optimizations of the local maximum skewness. When using unadjusted returns, the use of L-moments seems not to provide better results than with C-moments. In our experimentation, L-moment optimization is even worst than C-moments (the observed skewness is -.3 with C-moments versus -.57 with L-moments). Once again, we can argue that the L-moment is not reactive enough to take into account the distribution timing of the skewness. Nevertheless, when the returns are adjusted, the C-moment optimized portfolio is slightly improved in term of skewness (-.03 versus .01). The improvement is even more important when the adjusted returns are optimized with L-moments. This observation strengthens the idea of Kim and White (2004) that higher-order moments are mostly subject to bias. With no surprise, the performance measures show that L-moments, when applied to raw returns, failed to provide better portfolios. Nevertheless, optimizations done with ANN-GARCH corrected returns leads to a strong improvement of the $LMSP$, especially when L-moments are used. Again, we see that the ANN-GARCH correction leads to more concentrated allocation.

Furthemore, Table 3.10 shows the statistics on the different $LMKP$. Using unadjusted returns with L-moments or adjusted returns with C-moments leads to a comparable improvement. The combination of these two robust approaches allows us to obtain an even lower *kurtosis* portfolio. Nevertheless, this improvement of the fourth moments comes with

Table 3.8. Local Maximum Skewness Portfolios

	Moment 1	Moment 2	Moment 3	Moment 4
$LMSP$	10.10	18.85 [9.16]	-.03 [-.26]	21.16 [1.46]
LML_3P	9.43	19.56 [9.53]	-.57 [-.29]	18.60 [1.54]
$LM\breve{S}P$	13.85	18.09 [9.38]	.01 [-.11]	8.97 [1.36]
$LM\breve{L}_3P$	13.76	20.81 [10.60]	.21 [-.05]	10.82 [1.65]

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. The first two moments are annualized and expressed in %. The corresponding values of L-moments are indicated in brackets. L-moments 3 and 4 are multiplied by 1,000. $LMSP$ and LML_3P correspond to the Local Maximum Skewness and Local Maximum L-moment$_3$ Portfolios. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

Table 3.9. Performance Measures for Local Maximum Skewness Portfolios

	SR	CC	TR	WP_2	WP_{10}	WP_2^*	WP_{10}^*
$LMSP$	.37	6.84	1.72	.00	.00	.00	.00
LML_3P	.32	6.04	1.88	-.97	-2.44	-23.43	-26.56
$LM\breve{S}P$	.60	5.70	1.65	4.10	5.46	9.30	15.62
$LM\breve{L}_3P$	.51	4.22	2.16	2.83	.20	31.25	28.12

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. SR stands for the Sharpe Ratio, CC is for the Concentration Coefficient and TR denotes the Turn-over Ratio. The Willing-to-Pay measures are denoted WP_γ and WP_γ^* when they are computed respectively in the traditional Expected Utility and in the Cumulative Prospect Theory frameworks, γ being the level of the risk aversion. The corresponding values of the Willing-to-Pay measures (compared to the $LMSP$) are annualized and expressed in %. $LMSP$ and LML_3P correspond to the Local Maximum Skewness and Local Maximum L-moment$_3$ Portfolios. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

a weakening of the first two moments. Consequently, as shown in Table 3.11, performance measures that mostly focus on the first two moments are lower when a robust approach is chosen. But when we consider performance measures that focus on higher-order moments, robust portfolios are preferred.

Table 3.10. Local Minimum *Kurtosis* Portfolios

	Moment 1	Moment 2	Moment 3	Moment 4
$LMKP$	13.85	16.61 [8.63]	-.22 [-.17]	9.22 [1.29]
LML_4P	12.11	17.56 [9.15]	-.05 [-.17]	8.68 [1.32]
$LM\breve{K}P$	13.79	17.28 [9.05]	-.13 [-.15]	8.64 [1.26]
$LM\breve{L}_4P$	11.90	17.68 [9.28]	-.04 [-.14]	8.26 [1.26]

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. The first two moments are annualized and expressed in %. The corresponding values of L-moments are indicated in brackets. L-moments 3 and 4 are multiplied by 1,000. $LMKP$ and LML_4P correspond to the Local Minimum *Kurtosis* and Local Minimum L-moment$_4$ Portfolios. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

Table 3.11. Performance Measures for Local Minimum *Kurtosis* Portfolios

	SR	CC	TR	WP_2	WP_{10}	WP_2^*	WP_{10}^*
$LMKP$	.65	7.12	1.10	.00	.00	.00	.00
LML_4P	.51	6.49	1.86	-1.95	-3.22	9.23	6.25
$LM\breve{K}P$	.62	6.02	1.14	-.20	-1.17	4.68	3.12
$LM\breve{L}_4P$	.51	5.55	1.79	-2.24	-3.61	9.37	7.81

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. SR stands for the Sharpe Ratio, CC is for the Concentration Coefficient and TR denotes the Turn-over Ratio. The Willing-to-Pay measures are denoted WP_γ and WP_γ^* when they are computed respectively in the traditional Expected Utility and in the Cumulative Prospect Theory frameworks, γ being the level of the risk aversion. The corresponding values of the Willing-to-Pay measures (compared to the $LMKP$) are annualized and expressed in %. $LMKP$ and LML_4P correspond to the Local Maximum Skewness and Local Maximum L-moment$_3$ Portfolios. We add an extra $\breve{}$ onto the name of portfolios when they are obtained with corrected returns.

3.5 Conclusion

We propose in this paper an ANN-GARCH procedure to detect and correct outliers that occur in financial database. Based on the first work of Chen and Liu (1993) and Franses and Ghijsels (1999), Charles and Darné (2005) have developed a GARCH approach that detects and corrects Additive and Innovative Outliers. We extend this last work through the introduction of an Artificial Neural Network. The goal of this network is to capture a part of the non-linear relation that remains in unanticipated conditional volatility.

Using a French stock return database, we compare GARCH and ANN-GARCH outlier detections. Even if a ANN-GARCH model allows us to significantly reduce the unexplained conditional volatility, the detection of outliers provided by both models provides very similar results.

We then look for justification for the most significant detected outliers. A first quantitative approach allows us to differentiate global market shocks (systematic risk) from more stock-specific risks. A qualitative analysis based on past economic news complete the first quantitative analysis. Rational economic explanations for the detected outliers were found and in most cases linked to sound motives.

Then, we moved to the study of the impact of the outliers onto portfolio allocations. Using unadjusted and adjusted return series, we perform two kinds of optimizations: traditional C-moment and Robust L-moment optimizations. We evaluate the impact of detected outliers by considering the *ex post* moments of portfolios optimized and dynamically reallocated. Through our experiments, we first observe that the first, third and fourth moments are highly subject to biases when the data is not corrected for abnormal returns. Secondly, none of the robust methods improved the main characteristic of the Global Minimum Volatility Portfolio. Considering L-moments allows us to slightly reduce the impact of these outliers but the ANN-GARCH pre-processing seems more appropriate to limit the effects of aberrant returns. Our experiments suggest that a combination of these two robust methods leads to an improvement of the moment optimization. We also consider two complementary frameworks to evaluate the utility associated to each portfolio: a tra-

ditional power utility function (with various levels of risk aversion) in the Expected Utility framework and a "value" function in the context of the Cumulative Prospect Theory. Indeed, it is known that the Expected Utility framework may fail to efficiently capture the agent preferences for higher-order moments since it mainly focuses on the first two moments of return distributions. Both in the Expected Utility and the Cumulative Prospect Theory frameworks, our main over-all conclusion is that the ANN-GARCH-based method for deletion of outliers, coupled with the use of a robust approach based on higher-order L-moment, clearly improves the portfolio allocation process.

3.6 Appendices

3.6.1 Appendix 1: About Robust Estimation of GARCH(1,1) Models

We present here the traditional way to estimate GARCH models through Maximum Likelihood Estimation and Robust Bounded Maximum Likelihood Estimation proposed by Muler and Yohai (2008). For the sake of simplicity, we restrict the following estimations to the case of a GARCH(1,1) case.

We consider a series ε_t following a strong GARCH(1,1) process with $\varepsilon_t = \eta_t\sqrt{h_t}$ where the η_t are Independent and Identically Normally Distributed such as (with previous notations):

$$\begin{cases} \varepsilon_t = \eta_t\sqrt{h_t} \\ h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1\varepsilon_{t-1}^2. \end{cases} \tag{3.5}$$

From Equation 3.8, we have:

$$h_t = -\alpha_0\left(1-\beta_1\right)^{-1} + \sum_{k=1}^{t-1}\left(\alpha_1\beta_1^{k-1}\varepsilon_{t-k}^2\right), \tag{3.32}$$

Posing $\epsilon_t = \log(\varepsilon_t^2)$ and $\xi_t = \log\left(\eta_t^2\right)$, then we have $\epsilon_t = \xi_t + log\left(h_t\right)$. Since the Gaussian

density $f(.)$ is symmetric around 0, then the density $\phi(.)$ of ξ_t is given by:

$$\phi(\xi_t) = f[\exp(.5\xi_t)]\exp(.5\xi_t). \tag{3.33}$$

We consider $\theta = [\alpha_0\ \alpha_1\ \beta_1]$ as the set of parameters we want to estimate. From Equation 3.32, we obtained:

$$\hat{h}_t\left(\varepsilon_t^2|\theta\right) = -\alpha_0(1-\beta_1)^{-1} + \sum_{k=1}^{t-1}\left(\alpha_1\beta_1^{k-1}\varepsilon_{t-k}^2\right). \tag{3.34}$$

The usual form of the QML estimate based on the ε_t's consists of maximizing:

$$\begin{aligned}\hat{\theta} &= \underset{\theta\in\mathbb{R}^3}{\text{Argmax}} \left\{-.5\left\{\sum_{t=2}^{T}\varepsilon_t^2\left[\hat{h}_t\left(\varepsilon_t^2|\theta\right)\right]^{-1} + \sum_{t=2}^{T}\log\left[\hat{h}_t\left(\varepsilon_t^2|\theta\right)\right]\right\}\right\} \\ &= \underset{\theta\in\mathbb{R}^3}{\text{Argmax}} \left\{-\sum_{t=2}^{T}\exp\left\{\epsilon_t - \log\left[\hat{h}_t\left(\varepsilon_t^2|\theta\right)\right]\right\} \times \log\left[\hat{h}_t\left(\varepsilon_t^2|\theta\right)\right]\right\} \\ &= \underset{\theta\in\mathbb{R}^3}{\text{Argmax}} \left\{\sum_{t=2}^{T}\varphi\left\{\epsilon_t - \log\left[\hat{h}_t\left(\varepsilon_t^2|\theta\right)\right]\right\}\right\},\end{aligned} \tag{3.35}$$

where $\varphi(.) = -\log[\phi(.)]$.

The QML estimation is not robust since few outliers may have a large influence on recovered parameters (see also Zumbach, 2000). As pointed out by Muler and Yohai (2008), one reason for the lack of robustness of the QML-estimate is that $\varphi(.)$ is unbounded, and so large outliers may have an unbounded effect on $\hat{\theta}$. To gain robustness, Muler and Yohai (2008) proposed an M-estimate for GARCH parameters that includes a mechanism restricting the propagation of the outlier effect on estimated $\hat{h}_t$.

More precisely, the computation of the vector of parameters are provided by the estimation of the following equation for conditional volatility denoted $\breve{h}_{t:\iota}^*\left(\varepsilon_t^2|\theta,\iota\right)$, since depending on the parameter θ and on a threshold ι, with $\iota \in \mathbb{R}^+$, as such:

$$\breve{h}_{t:\iota}^*\left(\varepsilon_t^2|\theta,\iota\right) = \alpha_0 + \left[\alpha_1\chi_\iota\left(\overset{\cap}{\varepsilon}_{t-1}^2\right) + \beta_1\right]\breve{h}_{t-1:\iota}^*\left(\varepsilon_{t-1}^2|\theta,\iota\right), \tag{3.36}$$

where $\overset{\cap}{\varepsilon}_{t-1}^2 = \varepsilon_{t-1}^2 \times \left[\breve{h}_{t-1:\iota}^*\left(\varepsilon_{t-1}^2|\theta,\iota\right)\right]^{-1}$ is the reduced squared return series and with the

truncature function $\chi_\iota(.)$ depending on the arbitrary threshold ι, as such:

$$\chi_\iota(u) = u\mathbb{1}_{\{u\leq\iota\}} + \iota\mathbb{1}_{\{u>\iota\}}.$$

Finally, the Bounded QML-estimate of θ based on the ε_t consists of minimizing, like in the QML case (*Cf.* Equation 3.35), the following quantity (with previous notations):

$$\hat{\theta} = \underset{\theta\in\mathbb{R}^3}{\text{Argmax}} \sum_{t=2}^{T} \varphi\left\{\epsilon_t - \log\left[\breve{h}_{t:\iota}\left(\varepsilon_t^2\,|\theta,\iota\right)\right]\right\}. \tag{3.37}$$

Muler and Yohai (2008) show that the Bounded (Quasi) Maximum Likelihood estimator has both the properties of robustness against outliers and of consistency when the series follow GARCH models without outliers. From their Monte Carlo experimentations, they suggest using a value for the threshold ι equals to 4.

3.6.2 Appendix 2: Gaussianity Tests

Table 3.12. First Four Moments of Stocks of the Sample

	Annualized Mean			Annualized Volatility			Skewness			*Kurtosis*		
	Un-adjusted	GARCH Adjusted	ANN Adjusted	Un-adjusted	GARCH Adjusted	ANN Adjusted	Un-adjusted	GARCH Adjusted	ANN Adjusted	Un-adjusted	GARCH Adjusted	ANN Adjusted
1. ACCOR	9.36%	1.92%	2.45%	34.87%	31.63%	31.15%	.10	.01	.02	6.63	5.49	5.48
2. AIR LIQUIDE	8.99%	3.84%	4.20%	28.48%	28.20%	28.09%	.18	-.04	-.05	5.94	5.46	5.46
3. ALCATEL-LUCENT	-1.71%	-3.37%	-3.41%	54.22%	51.43%	52.64%	.17	-.33	-.12	16.98	7.17	6.66
4. AXA	8.58%	-2.43%	-2.39%	40.64%	35.11%	35.04%	.62	-.17	-.18	10.81	7.58	7.58
5. BNP PARIBAS	9.42%	3.04%	3.72%	37.24%	35.45%	35.00%	.14	-.10	-.11	8.75	6.28	6.28
6. BOUYGUES	17.32%	8.81%	9.22%	39.37%	36.75%	36.98%	.39	.06	.06	8.27	5.83	5.84
7. CAP GEMINI	14.04%	5.04%	5.56%	49.63%	47.89%	47.97%	.15	.00	.00	7.66	5.41	5.41
8. CARREFOUR	5.39%	.63%	1.57%	31.76%	30.57%	30.65%	.07	-.02	-.03	6.33	5.69	5.70
9. DANONE	10.94%	8.27%	9.15%	26.61%	25.75%	25.36%	.14	.07	.07	7.07	6.53	6.53
10. ESSILOR INTL	15.54%	8.69%	9.66%	30.50%	28.62%	28.95%	.59	.32	.31	8.23	6.71	6.70
11. L'OREAL	13.05%	5.95%	5.98%	31.86%	29.48%	28.99%	.22	-.03	-.04	6.38	6.01	6.02
12. LAFARGE	4.62%	-1.90%	-1.01%	34.39%	33.10%	33.42%	.06	-.19	-.20	6.50	5.61	5.61
13. LAGARDERE GROUPE	13.40%	8.01%	8.50%	38.51%	35.58%	35.92%	.54	.15	.14	9.36	5.92	5.93
14. LVMH	9.02%	2.56%	2.64%	34.82%	32.98%	33.45%	.44	.15	.16	7.38	5.26	5.25
15. MICHELIN	6.96%	-.78%	.05%	35.56%	35.25%	35.25%	.15	-.11	-.10	6.21	5.80	5.79
16. PERNOD-RICARD	14.28%	11.65%	11.02%	32.04%	30.45%	30.22%	.22	.27	.27	7.99	6.64	6.63
17. PEUGEOT	3.43%	-.24%	.36%	34.09%	31.49%	31.16%	.06	.10	.11	7.11	5.55	5.55
18. PPR	8.37%	2.57%	2.95%	36.34%	29.01%	28.94%	.41	.02	.02	8.46	7.42	7.43
19. RENAULT	5.24%	-.14%	-.10%	40.89%	37.50%	37.31%	.01	-.04	-.05	6.76	5.73	5.72
20. SAINT GOBAIN	8.70%	3.07%	4.05%	36.29%	32.61%	32.44%	.17	-.02	-.01	12.27	5.47	5.47
21. SANOFI-AVENTIS	15.82%	11.04%	11.82%	33.05%	32.21%	32.07%	.21	.11	.12	6.15	4.76	4.76
22. SCHNEIDER ELEC.	11.00%	7.87%	8.02%	36.91%	32.07%	31.77%	-.04	.00	-.01	8.39	5.21	5.22
23. SOCIETE GENERALE	9.98%	2.92%	3.57%	39.19%	37.66%	37.54%	.11	-.14	-.14	9.02	7.31	7.32
24. STMICROELEC.	10.43%	-2.79%	-2.46%	48.30%	47.81%	47.70%	.30	.06	.07	5.02	4.56	4.56
25. TOTAL	12.55%	7.17%	7.50%	30.37%	29.49%	29.14%	.19	-.08	-.09	7.20	5.49	5.49
26. UNIBAIL-RODAMCO	13.34%	10.55%	10.91%	25.90%	25.27%	25.03%	-.20	-.09	-.09	9.10	5.37	5.37
27. VALLOUREC	28.45%	18.71%	19.52%	40.82%	38.73%	38.59%	.39	-.01	.00	8.65	5.27	5.28
28. VINCI (EX SGE)	20.33%	13.09%	14.02%	33.65%	31.17%	31.54%	.63	.20	.20	9.43	5.35	5.35
29. VIVENDI	6.40%	3.79%	3.76%	39.46%	30.45%	35.25%	-.48	.05	.04	20.36	6.37	5.97

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors.

Table 3.13. Statistical Properties of Stock Returns

	J-B Stat.	J-B P-stat.	K-S Stat.	K-S P-stat.	A-D Stat.	A-D P-stat.
1. ACCOR	1826.42	.00	.06	.00	28.09	.00
2. AIR LIQUIDE	1211.40	.00	.06	.00	21.18	.00
3. ALCATEL-LUCENT	27011.15	.00	.08	.00	62.86	.00
4. AXA	8646.43	.00	.09	.00	72.34	.00
5. BNP PARIBAS	4577.39	.00	.08	.00	47.07	.00
6. BOUYGUES	3914.91	.00	.08	.00	44.28	.00
7. CAP GEMINI	3011.53	.00	.06	.00	84.73	.00
8. CARREFOUR	1533.02	.00	.06	.00	29.90	.00
9. DANONE	2299.70	.00	.07	.00	32.78	.00
10. ESSILOR INTL	3967.70	.00	.08	.00	43.22	.00
11. L'OREAL	1600.82	.00	.05	.00	19.66	.00
12. LAFARGE	1693.44	.00	.06	.00	27.49	.00
13. LAGARDERE GROUPE	5760.19	.00	.08	.00	65.09	.00
14. LVMH	2761.08	.00	.07	.00	32.46	.00
15. MICHELIN	1432.27	.00	.06	.00	29.24	.00
16. PERNOD-RICARD	3464.83	.00	.08	.00	47.42	.00
17. PEUGEOT	2338.10	.00	.06	.00	29.74	.00
18. PPR	4208.69	.00	.08	.00	50.31	.00
19. RENAULT	1957.72	.00	.06	.00	25.82	.00
20. SAINT GOBAIN	11890.55	.00	.07	.00	45.95	.00
21. SANOFI-AVENTIS	1397.99	.00	.06	.00	22.78	.00
22. SCHNEIDER ELECTRIC	4011.38	.00	.06	.00	25.99	.00
23. SOCIETE GENERALE	5020.29	.00	.08	.00	54.91	.00
24. ST MICRO.	616.95	.00	.06	.00	20.80	.00
25. TOTAL	2459.71	.00	.05	.00	20.91	.00
26. UNIBAIL-RODAMCO	5162.32	.00	.07	.00	34.16	.00
27. VALLOUREC	4492.47	.00	.07	.00	59.09	.00
28. VINCI (EX SGE)	5937.05	.00	.08	.00	81.16	.00
29. VIVENDI	41764.41	.00	.10	.00	76.14	.00

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. J-B Stat., J-B P-stat., K-S Stat., K-S P-stat., A-D Stat. and A-D P-stat. stand respectively for the Jarque-Bera Statistic, the Jarque-Bera P-statistic, the Kolmogorov-Smirnov Statistic, the Kolmogorov-Smirnov P-statistic, the Anderson-Darling Statistic and the Anderson-Darling P-statistic. P-statistics are expressed in %.

3.6.3 Appendix 3: Utility-based Performance Measures

Goetzmann and Kumar (2001) propose a portfolio diversification measure based on portfolio weights. It allows us for evaluating the naive diversification of a portfolio (without taking into account dependencies between asset returns). We propose here to consider the inverse measure: the Concentration Coefficient (denotes CC) defined as such:

$$CC = T^{-1} \sum_{t=1}^{T} \left[\sum_{n=1}^{N} \left(\mathbf{w}_{t,n}^2 \right) \right]^{-1} \tag{3.38}$$

Table 3.14. Heteroskedasticity Tests of Stock Returns

	McLeod-Li Test		LM Test	
	McL-L Stat.	P-stat.	Q-stat.	P-stat.
1. ACCOR	245.13	.00	244.85	.00
2. AIR LIQUIDE	92.68	.00	92.57	.00
3. ALCATEL-LUCENT	27.76	.00	27.72	.00
4. AXA	191.09	.00	190.86	.00
5. BNP PARIBAS	107.87	.00	107.74	.00
6. BOUYGUES	80.42	.00	80.37	.00
7. CAP GEMINI	38.40	.00	38.36	.00
8. CARREFOUR	189.13	.00	188.91	.00
9. DANONE	143.64	.00	143.47	.00
10. ESSILOR INTL	122.46	.00	122.32	.00
11. L'OREAL	178.85	.00	178.70	.00
12. LAFARGE	176.52	.00	176.44	.00
13. LAGARDERE GROUPE	43.65	.00	43.61	.00
14. LVMH	146.29	.00	146.12	.00
15. MICHELIN	81.58	.00	81.54	.00
16. PERNOD-RICARD	204.66	.00	204.43	.00
17. PEUGEOT	178.70	.00	179.41	.00
18. PPR	120.47	.00	120.33	.00
19. RENAULT	153.73	.00	153.98	.00
20. SAINT GOBAIN	44.29	.00	44.25	.00
21. SANOFI-AVENTIS	171.53	.00	171.33	.00
22. SCHNEIDER ELECTRIC	68.64	.00	68.56	.00
23. SOCIETE GENERALE	143.86	.00	143.91	.00
24. STMICROELECTRONICS (PAR)	50.47	.00	50.44	.00
25. TOTAL	248.56	.00	248.26	.00
26. UNIBAIL-RODAMCO	20.71	.00	20.68	.00
27. VALLOUREC	34.69	.00	34.65	.00
28. VINCI (EX SGE)	84.09	.00	83.99	.00
29. VIVENDI	345.83	.00	345.42	.00

Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. McL-L Stat. refers to the test of McLeod and Li (1983) for a GARCH(1,1) model estimated with Maximum Likelihood method, while the Q-stat. is the test statistic of the Lagragian Multiplier (*Cf.* Engle, 1982). The P-stat. correspond to P-statistics of related tests and are expessed in %.

where T and N are respectively the number of observations and assets, and $\mathbf{w}_{t,n}$ corresponds to the weights allocated to asset n at date t, with $t = [1, \ldots, T]$ and $n = [1, \ldots, N]$. This measure leads to an intuitive interpretation of portfolio concentration since CC varies from 1 for a low diversified portfolio (fully invested in one asset) to N for a highly diversified portfolio (the equally-weighted portfolio).

Another important measure in a dynamic setting is the allocation stability. It is traditionally measured with the Turnover Ratio (denoted TR with previous notations):

$$TR = (2T)^{-1} \sum_{t=1}^{T-1} \sum_{n=1}^{N} \left[|\mathbf{w}_{t+1,n} - \mathbf{w}_{t,n}| \right]. \tag{3.39}$$

The portfolio turnover measurement should be considered by an investor before deciding to purchase a given mutual fund or similar financial instrument. Indeed, a strategy with a high turnover rate will incur more transaction costs than another one with a lower rate. Unless the superior asset selection renders benefits that offset the added transaction costs they cause, a less active trading posture may generate higher returns. In addition, cost conscious fund investors should take note that the transactional brokerage fee costs are not included in the calculation of a fund's operating expense ratio and thus represent what can be, in high-turnover portfolios, a significant additional expense that reduces investment return.

As a complement measure, we use the satisfaction-based measure called Willing-to-Pay (West *et al.*, 1993; and Fleming *et al.*, 2001) to value the performance gains associated with a particular trading strategy. The idea is to estimate the performance fees an investor is ready to pay to switch between initial and targeted strategies (denoted as R and R^* respectively). Considering a general satisfaction function $u(.)$, the performance fee measure is defined as the sure return that has to be subtracted from the random return on the targeted strategy so that the investor becomes indifferent to both strategies:

$$u(R) = u(R^* - WP). \tag{3.40}$$

The satisfaction of the agent is evaluated in two different frameworks. First we choose to consider the traditional Expected Utility Theory framework and secondly, as a complement, the Cumulative Prospect Theory (*Cf.* Kahneman and Tversky, 1979).

Following Jondeau and Rockinger (2009), we choose to consider a Power-utility function

such as the Expected Utility written:

$$E\left[U\left(1+R\right)\right]=\left(1-\gamma\right)^{-1}\left(1+R\right)^{1-\gamma}, \tag{3.41}$$

with $\gamma > 1$ being the level of risk aversion.
Since our goal is to study the impact of the third and fourth moments on the investor preferences, we use a fourth-order Taylor series expansion (*Cf.* Jondeau and Rockinger, 2006) to approximate the level of utility provided by the investments, as such (with previous notations):

$$\begin{aligned} E\left[U\left(1+R\right)\right] \cong \; & \left(1-\gamma\right)^{-1}+m_1\left(R\right)-2^{-1}m_2\left(R\right)+6^{-1}\gamma\left(\gamma+1\right)\times \\ & m_3\left(R\right)-24^{-1}\gamma\left(\gamma+1\right)\left(\gamma+2\right)m_4\left(R\right), \end{aligned} \tag{3.42}$$

where $m_i\left(R\right)=E\left(R^i\right)$, for $i=[1,\ldots,4]$, are the non-central moments of order i. In this case, the Willing-to-Pay measure is obtained as follows:

$$E\left[U\left(1+R\right)\right]=E\left[U\left(1+R^*-WP\right)\right], \tag{3.43}$$

where $E\left(.\right)$ stands for the mean operator, $U\left(.\right)$ is the investor utility function, R and R^* are the initial and the targeted strategies and WP is the fix amount the investor is willing to pay to have R^* instead of R.

Secondly, since the Expected Utility theory still fails to explain some economic behaviors, we also propose to consider the Cumulative Prospect Theory introduced by Tversky and Kahneman (1979). As we already mentioned, the impact of higher-order moments in the Expected Utility framework seems to be quite insignificant in the portfolio choice of investors. Following Barberis and Huang (2007), we apply the Cumulative Prospect Theory in the context of portfolio choice. This last theory differs from the Expected Utility Theory in many points. For instance, the prospect theory focuses on gains and losses but not on the final wealth. The value function that drives investor preferences is concave over

gains but convex over losses. A subjective weighting function is applied to the Cumulative Distribution Function (see Figure 3.5). The effect of this transformation is mainly to make the agent overweigh the tails of the distribution.

We consider an investment with T possible returns represented by the complete set of the order statistics of returns $\left\{R_{[1:T]}, R_{[2:T]}, \ldots, R_{[T:T]}\right\}$, altogether with the associated Cumulative Distribution Function $F(.)$ and density $f(.)$, both related to the realizations of R.

Formally, the Cumulative Prospect Theory says that the agent evaluates an investment according to the associated satisfaction computed such as:

$$u(R) = \sum_{t=1}^{T} q_t v\left(R_{[t:T]}\right), \tag{3.44}$$

where $\forall t \in [1, \ldots, T]$:

$$q_t = \begin{cases} \varpi^+(f_t + \ldots + f_T) - \varpi^+(f_{t+1} + \ldots + f_T) & \text{for } R_{[t:T]} \geq 0 \\ \varpi^-(f_1 + \ldots + f_t) - \varpi^-(f_1 + \ldots + f_{t-1}) & \text{otherwise.} \end{cases}$$

and $v(.)$ is a value function which is concave over gains and convex over losses.

Following Kahneman and Tversky (1979), we use the following functional form (with previous notations):

$$v(R) = (R)^{\alpha} \mathbb{1}_{\{R \geq 0\}} - \lambda (-1 - R)^{\alpha} \mathbb{1}_{\{R < 0\}}, \tag{3.45}$$

where $v(.)$ is the value function, $\alpha \in [0, 1]$ and $\lambda > 1$ corresponding to the degree of sensitivity to risk and where the probability-weighting function is set as such:

$$\varpi^+[F(R)] = \varpi^-[F(R)] = \varpi[F(R)] \tag{3.46}$$

where:

$$\varpi[F(R)] = F(R)^{\delta} \left\{F(R)^{\delta} + [1 - F(R)]^{\delta}\right\}^{-1/\delta},$$

with $F(.)$ being the Cumulative Distribution Function associated to R.

It is to be noticed that several other probability-weighting functions have also been proposed in the lierature (*Cf.* for instance Wu and Gonzalez, 1996; Lattimore *et al.*, 1992; Prelec, 1998; Ingersoll, 2008):

$$\begin{cases} \varpi\left[F(R)\right] = F(R)^{\gamma}\left\{F(R)^{\gamma} + \left[1 - F(R)\right]^{\gamma}\right\}^{-\alpha} \\ \varpi\left[F(R)\right] = \alpha F(R)^{\gamma}\left\{\alpha F(R)^{\gamma} + \left[1 - F(R)\right]^{\gamma}\right\}^{-1} \\ \varpi\left[F(R)\right] = \gamma \exp\left\{-\beta\left[-\log F(R)\right]^{\alpha}\right\}, \end{cases} \tag{3.47}$$

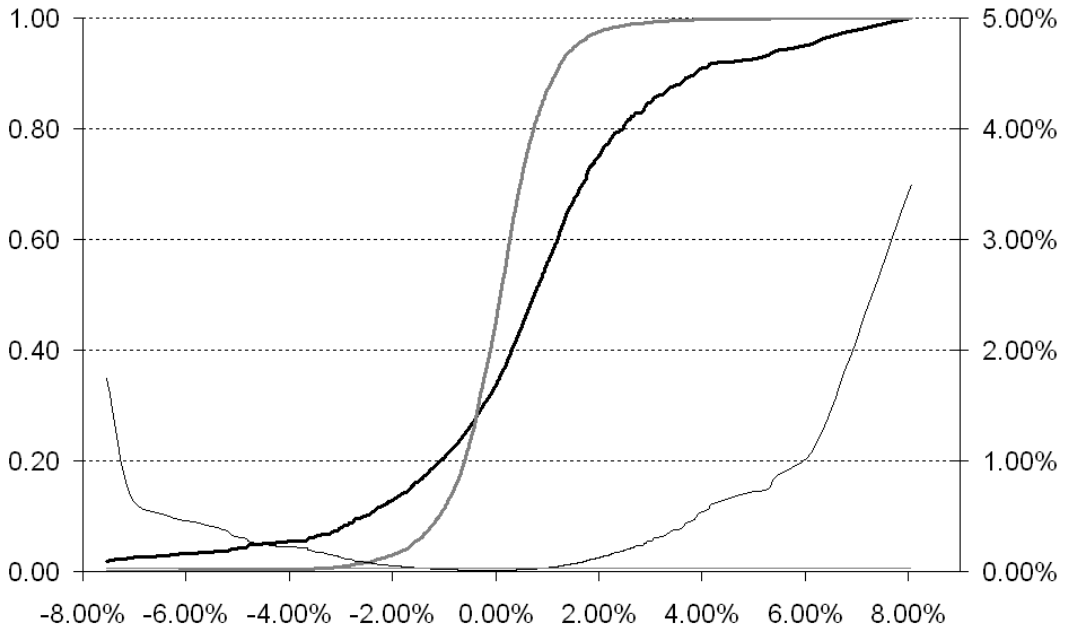
where $\gamma \in \left]\, 0, 1 \,\right]$, $\alpha > 0$ and $\beta > 0$.

In such Cumulative Prospect Theory framework, the revisited Willing-to-Pay measure, denoted here WP^{*}, is obtained as follows (with previous notations):

$$u(R) = \sum_{t=1}^{T} q_t v\left(R_{[t:T]}\right) = \sum_{t=1}^{T} q_t^{*} v\left(R_{[t:T]}^{*} - WP^{*}\right), \tag{3.48}$$

Some experimental evidences by Barberis and Huang (2007) suggest to use $\delta = .65$, $\alpha = .88$ and $\lambda = 2.25$.

Figure 3.5. Subjective *versus* Objective Probability Functions of the Returns on the Global Minimum Volatility Portfolio



Source: *Datastream*, daily quotes (01/01/1996-01/22/2009) in EUR; computations by the authors. This figure represents a specific probability-weighting function of returns on the Global Minimum Volatility Portfolio obtained with $\delta = .65$, $\alpha = .88$ and $\lambda = 2.25$. The black (grey) bold line represents the estimated kernel Transformed (raw) Cumulative Density Function of returns (on the left axis), whereas the black (grey) thin line represents the weighted (raw) Distribution (on the right axis).

Chapitre 4

Carte de style et facteurs de risque

4.1 Introduction

En quelques années, l'analyse de style s'est imposée comme référence incontournable de la gestion de portefeuille. Face à la grande diversité des types de gestion offerts, celle-ci est devenue fondamentale pour justifier des différences de performances entre les fonds d'un même univers. Une vaste littérature est consacrée à ce sujet. Les académiciens ont proposé plusieurs approches pour évaluer le style d'un fonds. Dans ce chapitre, nous supposons qu'un style est défini en fonction des facteurs de risque auxquels le fonds est exposé. Les sensibilités du fonds aux facteurs permettent d'évaluer les rendements futurs attendus et de vérifier la cohérence des performances passées par rapport aux objectifs définis. Pour les investisseurs, le style de gestion est ainsi une information primordiale.

Les modèles d'analyse de style ont évolué conjointement avec les modèles d'évaluation des actifs. Par exemple, les premières tentatives de regroupement des fonds utilisaient le risque systématique du « Modèle d'Évaluation des Actifs Financiers » (*CAPM*, *Cf.* Sharpe, 1964). Avec l'émergence de « l'*Asset Pricing Theory* » (*APT*) de Ross (1976), apparaît une nouvelle génération de modèles multifactoriels. Outre les retraitements statistiques possibles, les modèles de style vont se différencier à partir du choix des facteurs.

La définition des facteurs est très largement influencée par l'observation d'anomalies

de marché. De nombreuses études concluent à l'existence de biais dans les rendements des actions en fonction de la taille de la capitalisation (*Cf.* Lakonishok et Shapiro, 1986). D'autres (*Cf.* Lakonishok *et alii*, 1994 et Fama et French, 1998, par exemple) observent que le ratio de « valeur comptable sur cours »(*Book-to-Market*, noté B/M) est également déterminant lors de l'estimation du rendement d'un titre. Ces anomalies ont donné lieu à l'introduction de quatre facteurs explicatifs : les styles *Growth*, *Value*, *Large* et *Small*. Si les définitions précises des styles *Large* et *Small* sont immédiates, celles des styles *Growth* et *Value* sont plus incertaines. Le style *Growth* caractérise les entreprises dont on anticipe une croissance organique significative. Une stratégie d'investissement dite *Value* consiste à acheter des titres d'entreprises considérées comme sous-évaluées.

A l'échelle d'un portefeuille, deux grandes approches sont envisageables. La première consiste à déterminer les caractéristiques de style de chacun des titres composant ce portefeuille avant de les agréger (*Holding-based Style Analysis*). Les investisseurs ne disposant généralement pas de la composition des portefeuilles, la réalisation de ce type d'analyse de style est rarement possible. La seconde, introduite par Sharpe (1988 et 1992), permet de mesurer la sensibilité du portefeuille à différents facteurs de risque. Cette approche, communément appelée *Return-based Style Analysis* (RBSA) consiste en une régression contrainte des rendements du fonds sur ceux des facteurs de risque. Des résultats récents tendent à montrer qu'une approche de type *Holding-based Style Analysis* n'induit pas de meilleurs résultats que l'approche RBSA (*Cf.* ter Horst *et alii*, 2004). Ce modèle, bien que très largement répandu, est l'objet de sévères critiques. Le fait que les facteurs considérés soient généralement fortement colinéaires constitue l'une d'elles ; car alors, toute régression sur ceux-ci risque d'être fallacieuse.

Les modèles présentés jusqu'alors souffrent de deux limites majeures. L'analyse effectuée est statique ; les sensibilités du fonds aux facteurs sont supposées constantes sur l'ensemble de la période considérée. Pour un fonds géré passivement, cette approche est tout à fait valide. Mais dès lors que nous considérons des fonds adoptant une gestion active, celle-ci n'est plus viable. La seconde critique est liée à la nécessité de spécifier *a priori*

les facteurs de risque. Une connaissance préalable des fonds analysés est nécessaire afin de choisir les facteurs. Il faut ainsi connaître les biais susceptibles d'expliquer les évolutions du fonds. Ainsi Sharpe (1988) choisit des facteurs réels de l'économie (indices actions, obligations) sans pour autant intégrer les styles *Growth*, *Value*, *Large* et *Small*. Il faut attendre le modèle à trois facteurs pour que Fama et French (1998) proposent de prendre en considération ces derniers.

Brown et Goetzmann (1997) adoptent une approche originale. Exploitant le lien fort entre méthodes d'analyse factorielle et algorithmes de classification, ils proposent de réunir les fonds en groupes homogènes. A chacun de ces groupes correspond un facteur de risque (défini comme la moyenne des fonds de chaque groupe). Ce modèle, simple à mettre en œuvre, dispose d'un fort pouvoir explicatif. En particulier, les facteurs étant issus de l'univers de fonds, il n'est pas nécessaire de connaître *a priori* les styles. Les biais, même s'ils n'ont pas encore été observés, seront intégrés aux facteurs de risque.

A la suite de Maillet et Rousset (2003) et Aaron *et alii* (2004, 2005), nous choisissons les cartes auto-organisées de Kohonen (*Self-Organizing Maps*, SOM, *Cf.* Kohonen, 2000) pour développer un modèle d'analyse de style. Initialement présentées comme une méthode de classification, les SOM permettent la génération simultanée de facteurs de risque spécifiques et de groupes homogènes de fonds. En outre, les écueils liés à la colinéarité des facteurs sont évités. L'étape cruciale de sélection des indices de style représentatifs est de même écartée, les facteurs de risques intégrant implicitement les nouveaux biais potentiels. L'avantage des SOM est qu'elles autorisent une analyse dynamique. L'évaluation de la stabilité du style de gestion des fonds est rendue possible en projetant dynamiquement les fonds sur la carte. Nous appliquons notre méthode d'analyse de style à la population des fonds d'actions américaines présents dans la base *Lipper*. Elle regroupe les Valeurs Liquidatives (VL) de 598 fonds et 86 indices représentatifs du marché. La période de la classification s'étend du 30/08/2002 au 25/05/2007. Cette base a la particularité d'être renseignée avec le style communiqué par les gérants ; nous pouvons ainsi évaluer l'importance du biais de l'auto-déclaration. L'étude dynamique du style de gestion nous permet

de conclure à deux catégories de gestionnaires ; ceux adoptant une stratégie constante, fidèle à celle annoncée et ceux, plus opportunistes, qui changent régulièrement de style. Parmi ces derniers, nous différencions les gestionnaires anticipant efficacement les orientations futures du marché (les bulles) de ceux suivant simplement les tendances de marché.

Ce chapitre est organisé de la manière suivante. Dans la section 4.2, nous commençons par rappeler les principales méthodes RBSA. La section suivante est consacrée aux cartes de Kohonen, nous présentons la façon dont celles-ci sont utilisées pour déterminer les styles prépondérants des fonds. La section 4.4 illustre notre modèle en l'appliquant à un univers de fonds d'action américaine. La section 4.5 conclut.

4.2 L'analyse de style

Cochrane (2001) propose un modèle général qui réconcilie les deux grandes approches en matière de modèle d'évaluation des actifs : les modèles absolus (raisonnement à l'équilibre) et les modèles relatifs (raisonnement par arbitrage). Les modèles absolus sont définis à partir d'une économie à l'équilibre. Ils sont ainsi particulièrement adaptés à l'étude des interactions entre les actifs financiers et l'économie réelle. L'approche relative consiste à déterminer la valeur des actifs à partir de prix d'autres actifs, sans chercher l'origine de ces derniers. Reprenant sa formulation, nous présentons les différentes méthodes d'évaluation des actifs classiques avant de présenter les modèles d'analyse de style qui en découlent.

4.2.1 Les modèles d'évaluation

L'hypothèse principale du modèle général de Cochrane est que le prix d'un actif correspond à l'espérance de sa valeur future actualisée conditionnellement à l'information disponible :

$$P_t = E_t \left(m_{t+1} X_{t+1} \right), \tag{4.1}$$

avec P_t le prix de l'actif, $E_t(.)$ l'opérateur d'espérance conditionnelle à l'information disponible en t, X_{t+1} les valeurs futures de l'actif, et m_{t+1} le facteur d'actualisation stochastique.

Appliquée à l'actif sans risque, l'équation (4.1) nous permet d'établir que $E_t(m_{t+1}) = \frac{1}{R^f}$. Nous obtenons ainsi :

$$\begin{aligned} P_t = & \; E_t(m_{t+1})\, E_t(X_{t+1}) + Cov_t(m_{t+1}, X_{t+1}) \\ = & \; \frac{E_t(X_{t+1})}{R^f} + Cov_t(m_{t+1}, X_{t+1}). \end{aligned} \tag{4.2}$$

avec $Cov_t(.)$ l'opérateur de covariance conditionnelle à l'information disponible en t.

Nous retrouvons bien ici, la notion de prime de risque. Le rendement des actifs est égal au taux sans risque auquel s'ajoute une prime de risque. Cette prime de risque dépend directement du lien entre l'actif considéré et la consommation. Les actifs qui covarient avec la consommation sont ainsi plus volatils, la prime de risque associée sera donc importante. Au contraire, les actifs corrélés négativement avec la consommation (des assurances typiquement) vont offrir un rendement inférieur au taux sans risque. Nous réécrivons l'équation (4.2) en y appliquant l'espérance et en considérant les rendements (notés $R^i = \frac{X_t}{P_t}$) :

$$1 = E_t(m)\, E_t\left(R^i\right) + Cov_t\left(m, R^i\right). \tag{4.3}$$

Le *CAPM* (*Cf.* Sharpe, 1964) adopte un raisonnement d'équilibre permettant de définir un unique facteur de risque, le portefeuille de marché. Puisque $E_t(m_{t+1}) = \frac{1}{R^f}$, il vient :

$$E_t\left(R^i\right) = R^f - R^f Cov_t\left(m, R^i\right). \tag{4.4}$$

En considérant un facteur d'actualisation de la forme $m = a - bR^W$, avec R^W le portefeuille de marché (les valeurs de a et b sont obtenues par résolution d'un simple système), nous retrouvons ainsi la formulation habituelle du *CAPM* :

$$E_t\left(R^i\right) = R^f - \beta^i E_t\left(R^W - R^f\right), \tag{4.5}$$

avec $\beta^i = \frac{Cov_t\left(R^W, R^i\right)}{\sigma_t^2(R^W)}$ et $\sigma_t^2(.)$ l'opérateur de variance conditionnelle à l'information disponible en t.

L'*APT* proposé par Ross (1976) adopte une approche descriptive. Il suppose que les rendements des actifs sont influencés par différents facteurs exogènes (macroéconomiques). Le rendement attendu du portefeuille est fonction de ceux d'un panier de facteurs explicatifs et des sensibilités du fonds par rapport à ceux-ci. Dans le cadre du modèle général présenté par Cochrane, l'*APT* correspond au cas particulier où $m = a - b^{'} f$ avec $E(f) = 0$. L'introduction de ce facteur d'actualisation nous permet d'obtenir à partir de l'équation (4.3) :

$$1 = aE\left(R^i\right) - E\left(b^{'} f\right) E\left(R^i\right) + Cov\left(a, R^i\right) - Cov\left(b^{'} f, R^i\right). \tag{4.6}$$

Or $E\left(b^{'} f\right) = 0$ par construction et $Cov\left(b^{'} f, R^i\right) = E\left(b^{'} f \times R^i\right) - E\left(b^{'} f\right) E\left(R^i\right)$, ainsi :

$$E\left(R^i\right) = \frac{1}{a} - \frac{E\left(Rf'\right) b}{a}. \tag{4.7}$$

En posant $\beta^i = E\left(ff^{'}\right)^{-1} E\left(fR^i\right)$, $\alpha = \frac{1}{a}$ et $F = -\frac{E\left(ff^{'}\right) b}{a} = -\alpha E\left(mf\right)$, nous retrouvons la formulation générale de l'*APT* :

$$\mathbf{R}^i = \alpha + \beta^i \mathbf{F} + \varepsilon_i, \tag{4.8}$$

où $\mathbf{R}^i$ est le vecteur des rendements réalisés de l'actif i considéré, $\mathbf{F}$ est la matrice des rendements des facteurs de risque, β^i est la sensibilité de l'actif aux facteurs de risque, α est le rendement constant qui ne dépend pas des facteurs et ε_i est un bruit blanc.

4.2.2 Les modèles d'analyse de style

La spécification du facteur d'actualisation stochastique permet à Cochrane de retrouver les modèles traditionnels d'évaluation d'actif. De façon similaire, nous montrons dans cette partie comment les modèles d'analyse de style se distinguent essentiellement par le choix des facteurs.

Avec le *CAPM*, Sharpe (1964) adopte une approche statistique lui permettant de conclure à un unique facteur de risque, le portefeuille de marché. Il permet de relier le rendement attendu d'un titre à son risque systématique, au moyen d'une mesure d'exposi-

tion, le bêta. Jensen (1968) propose le premier regroupement d'actifs fondé sur un modèle d'évaluation d'actifs. Connor et Korajczyk (1986), Lehmann et Modest (1987) regroupent les actifs en fonction de leur risque systématique. Néanmoins, le *CAPM* souffre de nombreuses critiques (*Cf.* Roll, 1977). En particulier, deux anomalies en illustrent les limites. La première (*Cf.* Fama et French, 1998) provient de l'observation qu'un portefeuille de titres avec un faible ratio B/M est généralement plus rentable qu'un portefeuille ayant des titres de forts B/M. Les styles *Growth* et *Value* ont ainsi été définis. Le style *Growth* caractérise les entreprises démontrant une capacité à faire évoluer significativement leur taux de croissance et dont on anticipe une croissance forte. Elle dispose généralement d'un faible ratio B/M et d'un taux de distribution de dividendes très faible. Il s'agit typiquement de sociétés *leader* dans des secteurs modernes et dynamiques. Une entreprise est dite *Value* lorsqu'elle dispose d'un rendement dont l'évolution est régulière. Il s'agit souvent d'entreprises opérant dans des secteurs traditionnels avec une croissance de leurs revenus et profits lente mais régulière ; leurs ratios B/M sont élevés et leurs dividendes presque constants. Empiriquement, Lakonishok *et alii* (1994) observent que les firmes à fort B/M ont de faibles bénéfices, mais procurent des rendements moyens futurs supérieurs à ceux du marché. Inversement, les entreprises ayant un B/M faible, disposent de bénéfices élevés, mais des rendements futurs faibles.

La taille de la capitalisation est également un facteur discriminant. Les entreprises de petite capitalisation sont jugées plus risquées que celles disposant d'une large capitalisation. En outre, la liquidité est généralement moindre sur les titres des petites capitalisations. Il en résulte ainsi une prime de risque associée aux titres de faibles capitalisations.

La multiplication des travaux empiriques proposant d'autres variables explicatives des rentabilités excédentaires a rendu difficile la validation du *CAPM* (par exemple, Lakonishok et Shapiro, 1986 ; Chopra et Ritter, 1989 ; Fama et French, 1992). Les modèles d'évaluation ont évolué vers la prise en compte de facteurs de risque autres que le seul risque de marché.

En 1976, Ross propose l'*APT*. Il suppose que les rendements des actifs sont influencés

par différents facteurs exogènes (macroéconomiques). Le rendement attendu du portefeuille est fonction d'un panier de facteurs explicatifs et des sensibilités du fonds par rapport à ceux-ci. Pour être valide, ces modèles nécessitent d'avoir des facteurs indépendants. En effet, la colinéarité des facteurs peut conduire à une régression fallacieuse et ainsi induire un biais dans la détermination des sensibilités. Afin de s'accommoder de ces contraintes statistiques, des méthodes de décomposition factorielle ont été proposées. Ross utilise l'Analyse en Composantes Principales pour identifier des facteurs linéairement indépendants. Sous l'hypothèse de normalité des rendements, les composantes principales sont également indépendantes. La détermination de l'exposition du fonds aux différents facteurs orthogonaux (en évitant les écueils de multi colinéarité) est ainsi rendue possible. Soit $\mathbf{X}$ la matrice de taille $(K \times T)$ des rendements des indices de style choisis ; soit $\mathbf{P}$ la matrice de taille $(K \times T)$ des vecteurs propres de la matrice $\mathbf{XX'}$, avec $\mathbf{X'}$ la transposée de $\mathbf{X}$. La matrice des facteurs linéairement indépendants s'écrit alors $\mathbf{F} = \mathbf{XP}$. Pour que cette approche soit valide, il est nécessaire que les facteurs choisis soient Gaussiens. Une telle hypothèse est difficilement vérifiable en pratique. Une méthode récente de décomposition factorielle permet de s'affranchir de cette hypothèse : l'Analyse en Composante Indépendante (*ACI*, *Cf.* Hyvarinen *et alii*, 2001). Contrairement à l'Analyse en Composantes Principales, l'hypothèse de normalité des rendements est levée. L'*ACI* permet d'exprimer un ensemble d'observations multidimensionnelles comme une combinaison de variables latentes inconnues. Ces variables latentes inconnues sont appelées « les sources » ou « les composantes indépendantes ». Elles sont supposées être statistiquement indépendantes les unes des autres. L'indépendance est évaluée ici au-delà du moment d'ordre 2 et intégrée directement dans la fonction d'optimisation (*Cf.* Bonhomme et Robin, 2008). De façon formelle, en réécrivant sous forme matricielle la relation définie en (4.8), nous obtenons :

$$\mathbf{R} = \alpha + \mathbf{BF} + \mathbf{e} \Leftrightarrow \mathbf{F} = \mathbf{W'}(\mathbf{R} - \alpha - \mathbf{e}), \tag{4.9}$$

avec $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_N)$ la matrice de taille $(N \times T)$ des rendements des actifs, $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_K)$ la matrice de taille $(K \times T)$ des rendements des différents facteurs de risque, $\mathbf{B} = (\beta_1, \beta_2, ..., \beta_N)$ la matrice de taille $(N \times K)$ des sensibilités des actifs aux

facteurs de risque, α est le vecteur de taille (N) des rendements constants, $\mathbf{W}$ est la matrice de taille $(K \times N)$ d'extraction des facteurs, T la taille de l'échantillon et N le nombre d'actifs.

Pour avoir les facteurs, il suffit de trouver la matrice d'extraction. Le programme consiste en la maximisation de la non-Gaussianité des facteurs sous-jacents suivant la formule :

$$\mathbf{W} = \underset{\mathbf{W}}{\text{Argmax}}\, NG\left[\mathbf{W}(\mathbf{R} - \alpha)\right], \tag{4.10}$$

avec $NG(.)$ une fonction qui mesure la non Gaussianité des rendements.

Quelle que soit la méthode de décomposition factorielle choisie, l'APT fait face à deux limites principales. L'implémentation de ces modèles nécessite une connaissance *a priori* des facteurs auxquels est exposé le fonds considéré. Les facteurs de risque étant abstraits, l'interprétation des sensibilités du fonds à ces facteurs n'est pas toujours aisée.

Sharpe (1988 et 1992) présente un modèle adoptant cette fois une approche relative s'inscrivant dans le cadre de l'APT. Il considère que les paris et stratégies employés par le gérant (surpondération d'un type d'actif ou d'un secteur donné) vont définir le style du gérant. L'hypothèse du modèle de style de Sharpe est que les différences de comportement vont être répercutées directement sur les rendements des fonds gérés. Il choisi des facteurs réels de l'économie et considère un panier d'indices de marchés (actions, obligations...). Au moyen d'une régression multivariée sous contraintes, il détermine les sensibilités du fonds aux facteurs. En dépit du choix (arbitraire) des indices représentatifs et des problèmes d'ordre statistique liés à la colinéarité potentielle des facteurs utilisés, ce modèle devient une référence de l'analyse de style. En outre, l'utilisation de facteurs réels permet une interprétation directe des stratégies mises en œuvre par le gérant.

Fama et French (1996) proposent un modèle à trois facteurs. L'objectif du modèle est de pouvoir justifier les biais liés aux styles *Growth*, *Value*, *Large* et *Small*. Le premier facteur correspond à un facteur de marché classique. Les deux autres facteurs sont les facteurs « SMB » (correspondant à la différence des rendements d'un portefeuille de petites capitalisations et ceux d'un portefeuille de larges capitalisations) et « HML » (correspondant

à la différence des rendements d'un portefeuille ayant un ratio de B/M élevé moins ceux d'un portefeuille de B/M faible). L'intégration de ces nouveaux facteurs permet ainsi de palier (en partie) aux critiques des modèles précédents. De nouvelles anomalies de marché mettent en évidence les limites de ce modèle. De nouveaux facteurs explicatifs sont alors introduits. Cahart (1997) propose par exemple d'ajouter un quatrième facteur représentatif des stratégies dites « *momentum* », consistant à suivre les tendances de marché, les bulles spéculatives.

Les méthodes présentées sont purement statiques. Les bêtas obtenus sont figés sur la période d'analyse. Une telle approche est viable lorsque l'on considère un portefeuille géré passivement, mais va échouer si l'on évalue un gérant adoptant une gestion active. Les modèles précédents ne permettent pas de détecter les changements de stratégies employées par le gérant. Pourtant, dès 1996, Ferson et Schadt concluent à des changements de styles d'investissement en fonction des anticipations économiques. Il apparaît alors essentiel de prendre en compte les dynamiques des styles d'investissement (*Cf.* Chan *et alii*, 2002).

Brown et Goetzmann (1997) proposent une approche originale issue d'un algorithme de classification. L'idée consiste à créer des groupes homogènes de fonds. Une fois les groupes définis, les facteurs de risque sont simplement les moyennes des rendements des fonds des groupes. Les facteurs étant déterminés à partir des fonds eux-mêmes, ce modèle permet d'obtenir des facteurs de risque reflétant, entre autre, les stratégies actives de gestion.

Adoptant une approche similaire, Maillet et Rousset (2003) et Aaron *et alii* (2004) proposent d'analyser les styles des fonds en exploitant une méthode de classification appelée cartes auto-organisées de Kohonen (*Self-Organizing Maps*, notées SOM). Cet algorithme permet la détermination simultanée des groupes homogènes de fonds, et des facteurs de risque. Un avantage des cartes de Kohonen par rapport à l'algorithme utilisé par Brown et Goetzmann (1997), est la conservation de la topologie. Les facteurs représentatifs des groupes de fonds sont eux-même ordonnés sur la carte en fonction de leurs similitudes.

4.3 Les facteurs de style

L'algorithme des cartes de Kohonen permet d'obtenir des classifications non-linéaires sans connaissance *a priori* des données à classifier. Différentes applications financières exploitant les SOM ont été proposées : de la simple exploration d'univers de fonds, à la prévision de rendements futurs. Malgré sa robustesse, le caractère stochastique de l'algorithme peut engendrer des problèmes de convergence. Nous choisissons donc une version robuste des cartes de Kohonen (Guinot *et alii*, 2006). Nous présentons, dans cette section, l'algorithme SOM puis sa version robuste. Nous montrons ensuite, comment les cartes robustes peuvent être utilisées dans le cadre d'un modèle d'analyse de style.

4.3.1 Cartes auto-organisées de Kohonen

Une carte auto-organisée est un réseau constitué de neurones, unités ou vecteurs codes ou encore facteur de risque dans notre cas. Ils sont organisés sur une grille régulière de faible dimension. Nous considérons $I = [1, ..., K]$, l'ensemble des unités du réseau constitué de K neurones. Une structure de voisinage associée au réseau permet de définir la topologie de la carte.

L'état du réseau à l'instant l est donné par :

$$\mathbf{F}(l) = [\mathbf{F}_1(l), \mathbf{F}_2(l), ..., \mathbf{F}_K(l)], \tag{4.11}$$

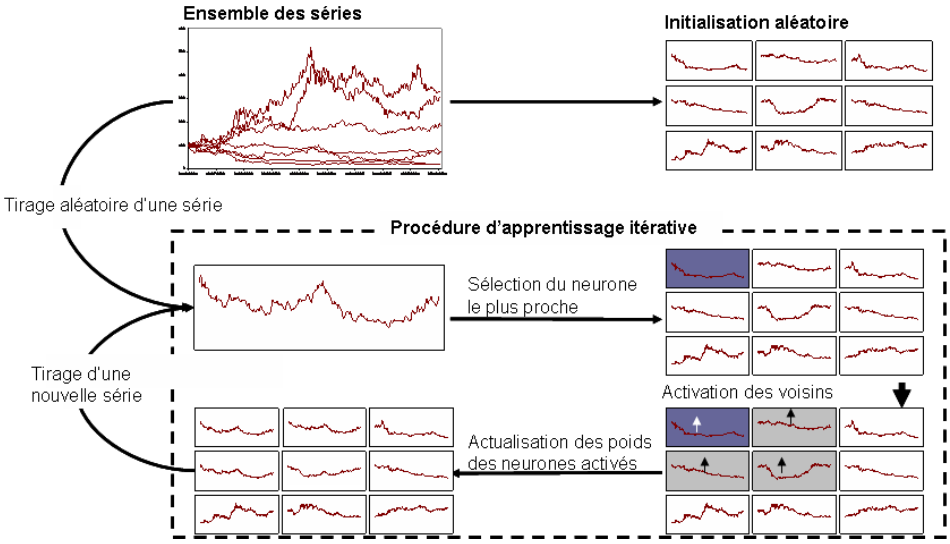
où $\mathbf{F}_k(l)$, pour $k = [1, ..., K]$ est un vecteur représentatif du facteur de risque et dont la dimension est égale à celle des données à classifier .

Etant donné un état du neurone $\mathbf{F}(l)$, et un input r_i, nous appellerons unité gagnante *BMU* (pour *Best Matching Unit*), le neurone dont le vecteur représentatif (nommé plus simplement le poids) $\mathbf{F}_{BMU_l}$ est le plus proche (en fonction de la métrique considérée, généralement la norme Euclidienne) de l'input $\mathbf{r}_i$.

L'algorithme SOM est tout d'abord initialisé de façon aléatoire dans l'espace des inputs. Puis, il est défini récursivement par les étapes suivantes :

1. tirer aléatoirement un input $\mathbf{r}_i$,

FIGURE 4.1. Représentation du processus itératif d'apprentissage de cartes auto-organisées.



2. trouver l'unité gagnante BMU tel que

$$BMU_{l+1} = \underset{k \in I}{\text{Argmin}} \left\{ \left\| \mathbf{r}_i \left(l+1 \right) - \mathbf{F}_k \left(l \right) \right\| \right\}, \tag{4.12}$$

où $\|.\|$ est généralement la norme Euclidienne,

3. Une fois le BMU trouvé, les poids des neurones sont actualisés de manière à ce que le BMU et ses voisins (les neurones activés par la fonction de voisinage) se rapprochent de l'input. La règle d'actualisation est la suivante :

$$\mathbf{F}_k \left(l+1 \right) = \mathbf{F}_k \left(l \right) - \varepsilon_l \Lambda \left[neight \left(\mathbf{F}_{BMU}, \mathbf{F}_k \right) \right] \left[\mathbf{F}_k \left(l \right) - \mathbf{r}_i \left(l+1 \right) \right], \tag{4.13}$$

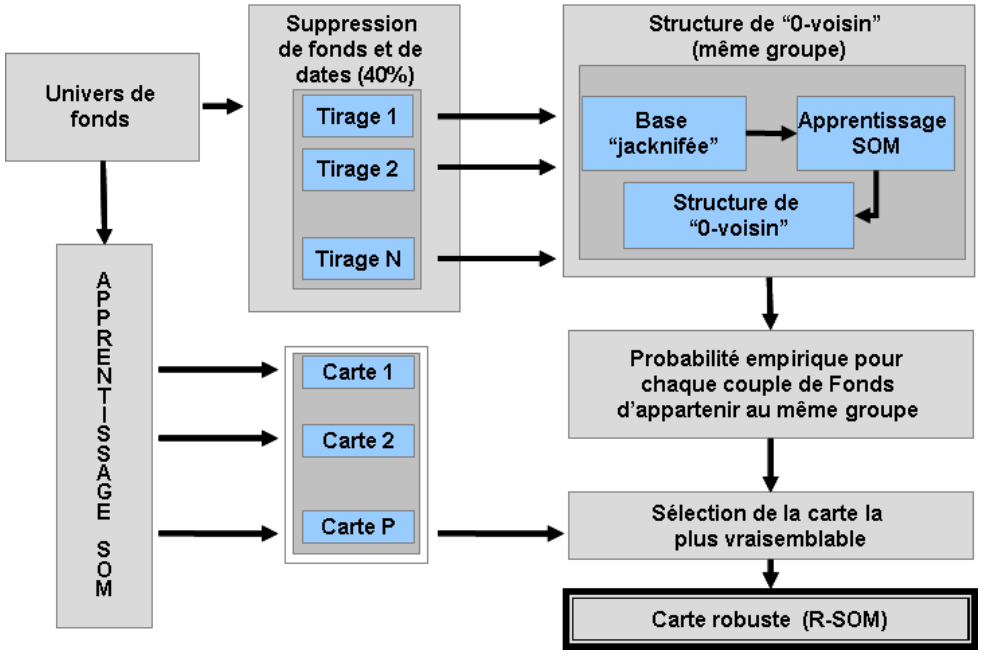
pour $k = [1, ..., K]$, avec $neight \left(., . \right)$ la fonction définissant le rang du voisinage entre deux neurones ($neight \left(\mathbf{F}_{BMU}, \mathbf{F}_k \right)$ vaut 1 si le BMU et $\mathbf{F}_k$ sont directement voisins, 2 si un neurone les sépare...) et où ε_l est le paramètre d'adaptation (décroissant avec le nombre d'itération et prenant des valeurs comprises entre]0,1[). Le nombre de neurones dont les poids sont actualisés durant la phase d'apprentissage dépend de la fonction de voisinage $\Lambda \left(. \right)$ qui est également décroissante avec le nombre d'itérations.

4.3.2 Cartes robustes

Si la qualité des classifications obtenues *via* SOM est généralement satisfaisante, le caractère stochastique de l'algorithme ne nous garantit pas une convergence systématique de la classification. Dès lors que deux classifications successives n'aboutissent pas aux mêmes résultats, utiliser une simple classification de Kohonen pour développer un modèle de style peut s'avérer hasardeux. Pour nous prémunir de tels problèmes de convergence, nous proposons d'appliquer une version modifiée de l'algorithme de Kohonen proposée par Guinot *et alii* (2006) : les cartes robustes (*Robust Self-Organizings Maps*, notées RSOM, *Cf.* Figure 4.2). Afin de limiter la dépendance du réseau aux données d'apprentissage du modèle, il est classique d'appliquer un processus de Bootstrap avec des techniques de ré-échantillonnage. L'idée de la technique du Bootstrap est ici appliquée à SOM en estimant une probabilité pour chaque individu d'être dans le même groupe. Cette probabilité est estimée empiriquement par le nombre de fois où deux individus sont « Zéro-voisins » (font partie du même groupe) lors de l'apprentissage SOM effectué sur les séries temporelles ré-échantillonnées. Par la suite, nous notons $\mathbf{P}$ la matrice contenant la probabilité empirique de deux individus d'être dans le même groupe à la fin d'une classification. L'algorithme de classification n'utilise que les individus présents dans la base ré-échantillonnée (60% des individus originaux). Nous généralisons cette approche en ajoutant un tirage sans remplacement des observations (60% des observations originales). A la fin de la première étape, les individus écartés lors du processus de ré-échantillonnage sont classifiés à partir de leurs distances aux vecteurs codes. Ainsi, à chaque étape de la classification, nous obtenons la probabilité d'être dans le même groupe pour chaque individu (y compris ceux écartés lors du ré-échantillonnage).

Lorsque la matrice $\mathbf{P}$ est construite, la première phase est terminée. Durant la seconde phase, l'algorithme SOM est à nouveau exécuté un grand nombre de fois (en conservant cette fois-ci l'ensemble des observations). A chaque carte obtenue M_i, nous construisons une table $\mathbf{P}_{M_i}$ similaire à la précédente (ses valeurs sont 1 si deux individus sont voisins et 0 sinon). La carte robuste sélectionnée (notée ci-dessous *RMap*) est celle qui minimise

FIGURE 4.2. Représentation de l'algorithme de cartes robustes.



la distance entre les deux structures de voisinage :

$$RMap = \underset{M_i, i \in I}{\text{Argmin}} \left\{ \|P - P_{M_i}\|_{Frob} \right\}, \tag{4.14}$$

avec $\|A\|_{Frob} = \frac{1}{n^2} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{[i,j]}^2}$, et n la dimension de la matrice carré $\mathbf{A}$, dont les éléments sont $a_{[i,j]}, \forall (i,j) \in I^2$.

4.3.3 Analyse de style par cartes robustes

Nous abordons maintenant la présentation du modèle d'analyse de style fondé sur les cartes robustes. L'objectif est de définir un modèle multifactoriel qui prenne en compte les dynamiques de gestion et qui permette d'éviter les biais liés au choix des indices de style représentatifs. Un moyen d'y parvenir est d'extraire les facteurs de façon non supervisée de l'univers des fonds. Ainsi, à partir des performances passées des fonds, nous effectuons une classification en carte robuste. Cette carte nous permet d'obtenir simultanément des groupes de fonds homogènes ainsi que des facteurs représentatifs de chaque groupe. Les

vecteurs codes caractéristiques des classes obtenues sont alors considérés somme les facteurs de style de notre modèle. L'extraction non supervisée de ces facteurs de risque nous permet d'éviter de devoir spécifier *a priori* les facteurs de style représentatifs. L'ensemble des biais de gestion adoptés par les gérants est, en outre, intégré par ces facteurs. Si, au sein de notre univers d'analyse, une part significative de gérants adopte des changements de gestion, les facteurs obtenus par la carte robuste intègrerons cette dynamique.

Notre modèle d'analyse de style est entièrement défini par les groupes obtenus. Il est ainsi naturel de s'interroger sur la calibration optimale de la carte. Il nous faut spécifier deux paramètres principaux, la nature des données considérées pour prendre en compte les performances passées (valeurs liquidatives, rendements, rendements relatifs...) et la taille de la carte. Dans une étude préliminaire (*Cf.* Annexe 4.6.2), nous rappelons les critères standard permettant de juger de la qualité des cartes de Kohonen. Nous considérons spécifiquement les erreurs topologiques et de quantification ainsi que deux critères *ad hoc*, les corrélations *intra* et *extra* classes. Les éléments principaux de calibration sont : la taille de la carte, les prétraitements possibles de la base (VL brutes, VL relatives, rendements et rendements relatifs), le choix des fonctions de voisinage et de distances intervenant lors de l'apprentissage des SOM (considérées aux équations 4.12 et 4.13 respectivement). A l'aune de ces critères, nous établissons que les paramètres les plus adaptés à notre problème de classification de fonds correspondent à une carte de dimension (4×4) créée à partir des rendements relatifs des fonds et dont l'apprentissage est effectué avec une fonction de voisinage Gaussienne et une fonction de distance Euclidienne.

Il nous faut ensuite traiter le problème de l'interprétation des facteurs obtenus. Une première méthode consiste à projeter des indices (de style ou sectoriels) connus sur la carte. Leur classe d'appartenance sera ainsi représentative du style correspondant. Du fait des fortes colinéarités des indices de styles entre eux, cette première approche n'est pas toujours satisfaisante, elle ne permet pas nécessairement d'identifier avec précision les styles de chacun des groupe. Nous en proposons donc une seconde, basée sur une classification conditionnelle des fonds et des indices. L'hypothèse de Sharpe est alors reprise : le gestionnaire ayant retenu un style particulier doit surperformer lorsque le facteur cor-

respondant surperforme. Nous considérons, pour chaque indice standard, les périodes sur lesquelles celui-ci domine le marché. Ces périodes sont retenues pour réaliser des classifications conditionnelles. Les nouvelles classes conditionnelles sont alors indexées en fonction de la performance des individus représentatifs de chacune. A chaque fonds est attribué le score de la classe conditionnelle à laquelle il appartient. Le style des groupes inconditionnels est alors déterminé en considérant les scores conditionnels de l'ensemble des fonds qu'elle regroupe.

Les facteurs explicités, la dynamique des styles des fonds est alors évaluée. Pour cela, nous retenons un ensemble de périodes pour lesquelles nous souhaitons évaluer ces styles. Pour chacune de ces sous-périodes, nous déterminons le positionnement des fonds sur la carte (au moyen de la projection des séries de performances des fonds restreintes à la période considérée sur les vecteurs codes, en appliquant l'équation 4.12).

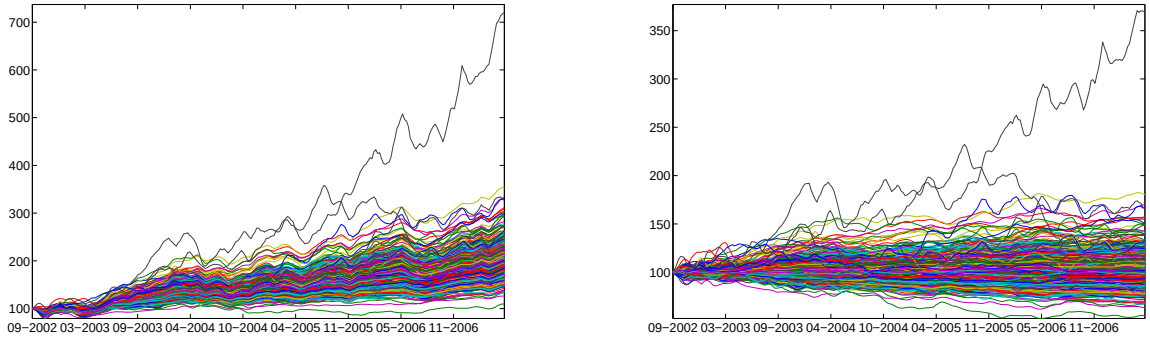
L'étude du positionnement dynamique des fonds sur la carte, nous permet d'évaluer la stabilité des styles retenus par les différents gestionnaires. Parmi les gérants jugées instables, nous cherchons à différencier ceux anticipant avec succès les futures conditions de marché (*market timer*), de ceux adoptant simplement une approche *momentum*, c'est-à-dire, suivant les tendances de marché, les bulles spéculatives.

Nous introduisons pour cela la notion de « pari actif du gérant ». Chaque changement de gestion constaté (matérialisé par une migration sur la carte) va correspondre à un pari. Il s'agit donc de déterminer si ces paris sont gagnants ou pas. A chaque changement de classe de la carte, nous considérons les deux vecteurs codes représentatifs des classes (avant et après changement). Nous comparons leurs performances sur la sous-période consécutive au changement de classe. Si la performance du vecteur code de la classe de départ est inférieure à celle du vecteur code de la classe d'arrivée, alors nous supposerons que le gérant a gagné son pari. Le nombre de paris gagnants rapporté au nombre de paris total nous renseigne sur la capacité du gérant à anticiper les futures conditions de marché.

4.4 Détermination des styles par cartes robustes sur le marché des fonds d'actions américaines

Nous choisissons d'illustrer notre modèle d'analyse de style par l'étude de la population des fonds actions américains. Les données sont fournies par *Lipper*, la base est constituée des Valeurs Liquidatives (VL) en dollar de 598 fonds (*Cf.* Figure 4.3). La période de l'étude s'étend du 30 août 2002 au 25 mai 2007. La fréquence d'observation des VL est hebdomadaire (cotation le vendredi).

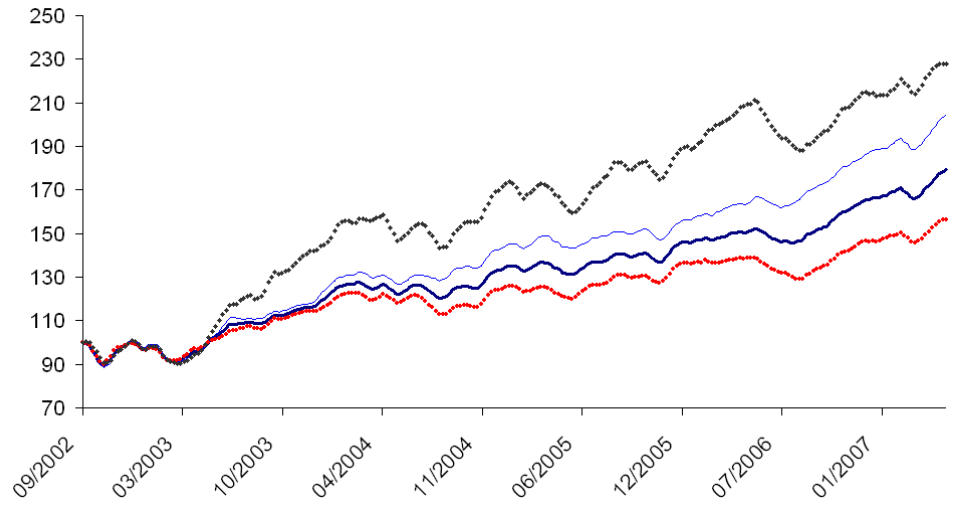
FIGURE 4.3. Evolution des performances des fonds du marché (absolues à gauche, relatives à droite.



Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les valeurs liquidatives absolues et relatives (par rapport à la moyenne des fonds) représentées ci-dessus sont basées à 100 en début de période.

Afin de faciliter l'interprétation des cartes obtenues, nous incluons à l'univers de fonds, un large ensemble d'indices. Nous disposons au total de 86 indices représentatifs du marché. Certains rendent compte des performances du marché global alors que d'autres, plus spécialisés, illustrent l'évolution de secteurs ou de styles bien précis (*Cf.* Figure 4.4). Cette représentation nous permet également d'illustrer les problèmes potentiels auxquels font face les modèles d'analyse de style traditionnels. En effet, durant les 9 premiers mois, les 4 indices de styles retenus (*MSCI USA*, *MSCI USA Value*, *MSCI USA Growth* et *MSCI USA Small Cap*) sont parfaitement confondus. Sur les périodes suivantes, les différences de style sont plus marquées, mais les indices restent visiblement très dépendants.

FIGURE 4.4. Evolution des performances des indices de style du marché.



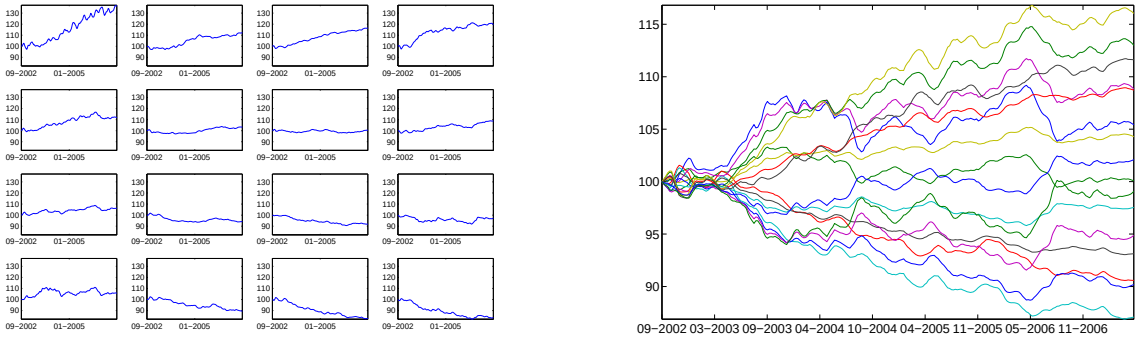
Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006; calcul des auteurs. Les valeurs liquidatives absolues représentées ci-dessus sont basées à 100 en début de période. Les courbes épaisse, fines, pointillée et tiret représentent respectivement l'évolution du *MSCI USA*, *MSCI USA Value*, *MSCI USA Growth* et *MSCI USA Small Cap*.

4.4.1 Création et validation de la carte

Nous classifions *via* l'algorithme des cartes robustes, la base des fonds et indices américains à partir de leurs rendements relatifs. Les performances représentées sur la Figure 4.5 correspondent aux « vecteurs codes » des différentes classes. Ces classes, ainsi que les facteurs de risque correspondants ont été définis par l'algorithme SOM lui-même. Afin de pouvoir se référer à une cellule particulière, nous numérotons de 1 à 16 (dans une carte 4×4) chacune d'elle. Par convention, la numérotation se fait de bas en haut (axe Sud-Nord par analogie avec une carte géographique) puis de gauche à droite (axe Ouest-Est), la case 1 est donc la case en bas à gauche et la case 16 en haut à droite (les cases 4 et 13 sont donc, respectivement, les angles Nord Ouest et Sud Est).

Une première étape concernant la validation de la carte consiste à évaluer les similitudes entre chacune des classes et à s'assurer du bon ordonnancement de ces dernières. La matrice des distances unifiées (U-matrice, *Unified distance matrix*) permet la représentation en deux dimensions des distances entre les unités de la carte (*Cf.* Figure 4.6). Plus formellement, si la carte initiale est de taille $(n \times m)$, alors la U-matrice est de taille

FIGURE 4.5. Performances relatives représentatives des seize classes (pour chaque cluster à gauche, superposées à droite).



Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les valeurs liquidatives relatives (par rapport à la moyenne des fonds) représentées ci-dessus sont basées à 100 en début de période.

$[(n+n-1)\times(m+m-1)]$ et est définie par :

$$\mathbf{U}(i,j)=\begin{cases}\left\|\mathbf{F}\left(\frac{i+1}{2},\frac{j}{2}-1\right)-\mathbf{F}\left(\frac{i+1}{2},\frac{j}{2}+1\right)\right\| \text{ si } i \text{ est impaire et } j \text{ est paire,}\\ \left\|\mathbf{F}\left(\frac{i}{2}-1,\frac{j+1}{2}\right)-\mathbf{F}\left(\frac{i}{2}+1,\frac{j+1}{2}\right)\right\| \text{ si } i \text{ est paire et } j \text{ est impaire,}\\ \text{la moyenne des } \mathbf{U}(k,l) \text{ adjacents à } \mathbf{U}(i,j) \text{ sinon.}\end{cases} \tag{4.15}$$

Nous vérifions ainsi que les proximités des classes sur la carte respectent celles des individus dans la base, et que la base de données ne correspond pas à une topologie trop particulière pour être prise en compte par l'algorithme (« trou au milieu de la carte », forme dite « en fer à cheval »...). Au regard de la Figure 4.6, nous constatons que la carte se scinde en deux zones, la partie basse et la partie haute. Cette césure se matérialise par la bande blanche horizontale au milieu de la figure. Ces deux groupes se subdivisent (moins nettement) ensuite en deux parties (est et ouest). De façon générale, les distances observées sont faibles. La carte semble ainsi bien représenter le *continuum* des profils des fonds. Il est donc possible d'interpréter les proximités des cellules de la carte en termes de similarités, sans qu'il soit vraiment nécessaire de prendre de plus amples précautions d'analyse.

L'organisation et la cohérence globale de la carte validées, nous nous intéressons mainte-

FIGURE 4.6. Carte des distances séparant les individus représentatifs des classes.

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les valeurs liquidatives absolues représentées ci-dessus sont basées à 100 en début de période. Le graphique représente les distances deux à deux, des individus représentatifs concomittant.

nant à l'homogénéité des différents groupes. Les R^2 *intra* classes sont obtenus en régressant les rendements (bruts) des fonds d'une classe sur ceux du meilleur indice de la classe, défini comme celui, parmi ceux rangés dans la dite classe, qui donne le meilleur R^2, en moyenne, pour l'ensemble des fonds de la classe considérée. Les autres coefficients de détermination (R^2 *extra* classe et R^2 population) sont issus de la régression des rentabilités brutes des fonds à l'extérieur de la classe considérée, sur ceux de l'indice le plus représentatif de cette classe. Un coefficient de détermination moyen R^2 *intra* classe supérieur au R^2 moyen *extra* classe valide le fait que l'indice retenu est bien caractéristique du style de gestion des fonds de la classe considérée. Les coefficients de corrélation de Pearson moyens (notés dans la table 4.1, respectivement, C moyen *intra* classe et C moyen *extra* classe) sont obtenus en moyennant les coefficients de corrélation linéaire des rentabilités relatives (à la moyenne générale) des fonds, calculés deux à deux (respectivement, par paire d'individus à l'intérieur d'une même classe et par paire d'individus dont l'un est à l'intérieur et l'autre à l'extérieur de cette classe).

Les coefficients *intra* classes, toujours plus élevés que ceux *extra* classes, peuvent s'interpréter comme des indicateurs de la cohésion interne des classes – au sens du coefficient

TABLE 4.1. Caractéristiques linéaires de la classification.

Classe	Nombre de fonds	Nombre d'indices	R^2 moyen intra classe	R^2 moyen extra classe	R^2 moyen population entière	C moyenne intra classe	C moyenne extra classe
1	100	18	0.94	0.78	0.80	0.65	-0.05
2	34	8	0.94	0.81	0.82	0.50	0.02
3	56	7	0.93	0.82	0.83	0.55	-0.01
4	41	2	0.91	0.80	0.80	0.46	-0.02
5	34	3	0.92	0.80	0.80	0.48	0.02
6	13	1	0.90	0.84	0.84	0.08	0.01
7	11	2	0.82	0.84	0.84	0.13	0.00
8	39	8	0.84	0.74	0.74	0.37	-0.05
9	36	1	0.87	0.85	0.85	0.26	0.00
10	30	3	0.91	0.85	0.85	0.17	-0.02
11	8	2	0.93	0.85	0.85	0.15	-0.02
12	34	4	0.87	0.80	0.80	0.45	-0.07
13	18	8	0.82	0.85	0.85	0.32	-0.01
14	43	7	0.86	0.85	0.85	0.41	-0.06
15	30	2	0.89	0.85	0.85	0.40	-0.06
16	72	10	0.92	0.79	0.80	0.57	-0.12

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les coefficients R^2 sont issus des régressions des rentabilités hebdomadaires brutes sur celles de l'indice (donnant les meilleurs résultats moyens par classe) ; les coefficients C sont les corrélations des rentabilités relatives par rapport à la moyenne des fonds pris deux à deux, et considérés à l'intérieur d'une même classe (*intra*), à l'intérieur et à l'extérieur (*extra*).

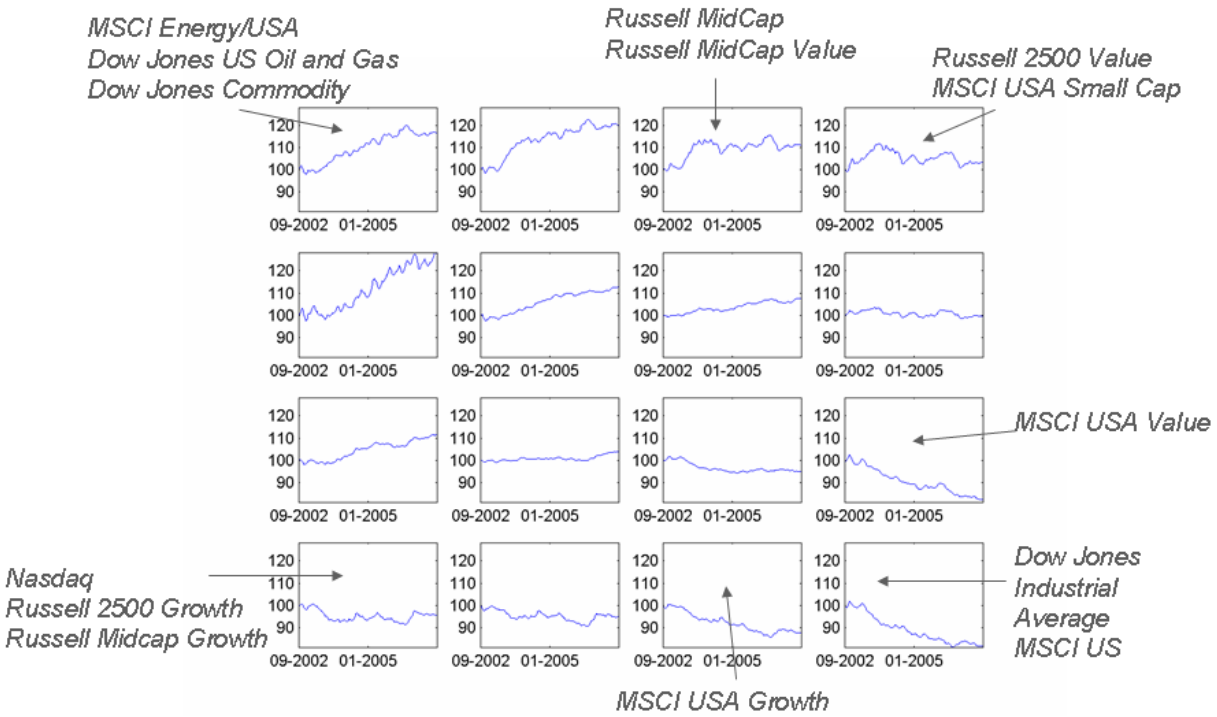
de corrélation linaire de Pearson. S'agissant de performances relatives, le fait remarquable ici est que les coefficients *intra* classes sont en moyenne 7 fois plus importants que les coefficients *extra* classes. Ainsi, même si les corrélations linéaires ne sont pas prises en compte directement dans l'algorithme, la classification obtenue (qui s'attache directement à l'évolution des performances des fonds) met en évidence les liens qui existent entre les évolutions des performances de fonds et celles d'indices de style.

4.4.2 Interprétation de la carte des fonds en terme de styles : approches classique et conditionnelle

Nous abordons la recherche des biais de style. Une première idée naturelle est de repérer les indices de style au sein de la carte. Les analyses précédentes nous ont permis de valider la cohérence de la carte (en terme de proximité des classes sur la carte et des individus dans la base, ainsi qu'en terme de topologie), nous pouvons ainsi exploiter la propriété de conservation de la topologie de l'algorithme RSOM pour déduire les styles prépondérant de chacune des classes. Parmi les 86 indices présents dans notre base, nous en retenons

4 comme représentatifs des biais principaux : le *MSCI USA* pour le biais *Large*, *MSCI USA Growth* pour le biais *Growth*, *MSCI USA Small Cap* pour le biais *Small*, *MSCI USA Value* pour le biais *Value*. Nous incluons également l'indice global *Dow Jones Industrial Average* pour déterminer les fonds indiciels ou sans biais ainsi que le Nasdaq pour les fonds à dominante technologique. La Figure 4.7 indique le positionnement de ces indices sur la carte (obtenu par la détermination du vecteur code le plus similaire, *Cf.* équation 4.12).

FIGURE 4.7. Repérage des indices représentatifs et des facteurs sur la carte.



Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les valeurs liquidatives relatives représentées ci-dessus correspondent aux rendements relatifs des individus représentatifs (vecteurs codes).

Cette première tentative de repérage des styles est en ligne avec l'analyse issue des corrélations *intra* classes ; les différents biais de style se retrouvent représentés au sein de classes situées aux extrémités de la carte. Nous identifions quatre styles prépondérants : les fonds « matières premières » et « énergies » (en haut à gauche), les fonds indiciels globaux (en bas à droite), les fonds *Small Cap* et *Mid Cap* (en haut à droite) et les fonds « technologiques »et *Small Growth* (en bas à gauche).

L'analyse précédente, trop sommaire, ne nous permet pas de d'identifier avec précision

les biais de gestion recherchés. Les indices de styles technologique et *Mid Growth*, en particulier, se retrouvent affectés à la même classe (en bas à gauche de la carte).

TABLE 4.2. Matrice de corrélation des indices de style.

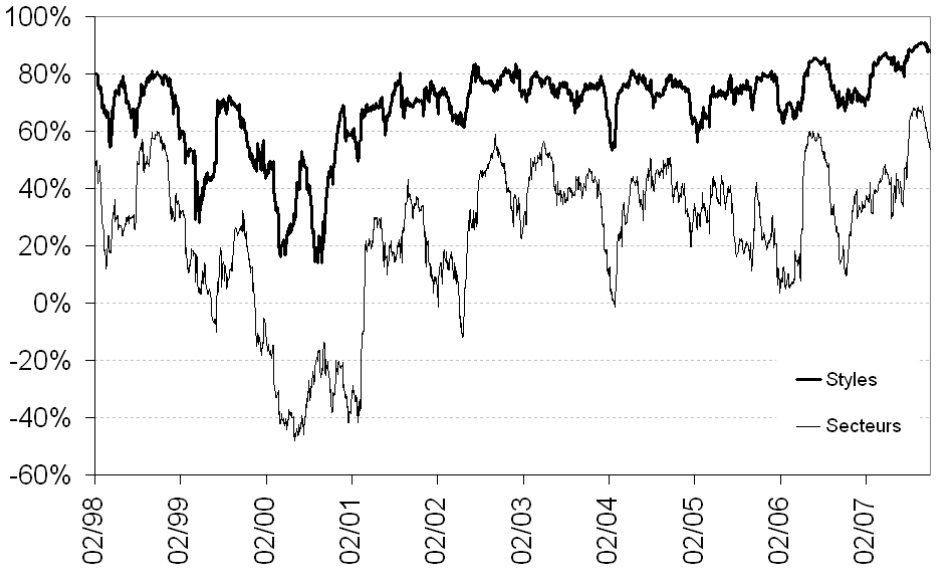
	Dow Jones Industrial Average CR	*MSCI USA TR USD*	*NASDAQ Composite CR*	*MSCI USA Growth TR USD*	*MSCI USA Value TR USD*	*MSCI USA Small Cap TR USD*
DJ Industrial Average CR	1	0.96	0.83	0.92	0.94	0.80
MSCI USA TR USD	0.95	1	0.90	0.96	0.97	0.88
NASDAQ Composite	0.81	0.89	1	0.93	0.83	0.90
MSCI USA Growth USD	0.89	0.96	0.93	1	0.88	0.86
MSCI USA Value USD	0.93	0.95	0.78	0.84	1	0.85
MSCI USA Small Cap	0.79	0.88	0.89	0.86	0.83	1

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Cette table présente les corrélations de Pearson (sur la partie triangulaire supérieure) et les corrélations de Spearman (sur la partie triangulaire inférieure) des séries de rendements des indices de style considérés.

La table 4.2 présente les corrélations de Pearson et de Spearman des rendements des différents indices de style considérés, respectivement sur la partie triangulaire supérieure et triangulaire inférieure. Les *P-values* correspondantes sont toutes très inférieures à 0.001 ; les corrélations présentées sont ainsi significatives. L'indice des fonds *Small Cap*, le *MSCI USA Small Cap TR* a la plus faible corrélation avec les autres indices (entre 0.79 et 0.94). Les corrélations entre les indices *Value* et *Growth*, sont également assez faibles (à 0.88 et 0.84 respectivement pour les corrélation de Pearson et Spearman). Le Nasdaq présente de faibles corrélations avec les indices *Larges* (en particulier ses corrélations de Pearson et de Spearman avec le Dow Jones Industrial Average CR sont respectivement de 0.83 et 0.81) et *Value* (0.83 et 0.78). Il présente, au contraire, une dépendance élevée vis-à-vis du style *Growth* (0.93). Les autres mesures de corrélation traduisent de fortes dépendances des indices entre-eux. Les indices *Global*, *Large*, *Growth* et *Value* ont des corrélations toutes supérieures à 0.88. Les très fortes dépendances linéaires entre ces indices nous confortent donc dans l'utilisation de classification non-linéaire pour déterminer les styles des fonds. La dynamique des dépendances des facteurs de style entre eux est, de plus, fortement instable. La Figure 4.8 présente l'évolution des corrélations de Pearson des indices *DJ Stoxx Growth*

et *DJ Stoxx Value* ainsi que celle des indices sectoriels *DJ Stoxx*. Les corrélations des indices de style fluctuent de moins de 20% à plus de 80%. Celles des indices sectoriels sont encore plus instables (en moyenne, les indices sont positivement corrélés à 0.3, mais ils peuvent atteindre des niveaux de décorrélations importants). Les caractéristiques des dépendances linéaires (élevées et instables) nous incitent ainsi à l'adoption d'une analyse conditionnelle et non-linéaire afin de déterminer les niveaux d'exposition des fonds aux styles.

FIGURE 4.8. Evolution des corrélations des indices de style *Large* et *Small*, et *Growth* et *Value*.



Source : *Datastream*, base d'indices européens en euros, données quotidienne du 12/11/1998 au 12/11/2007 ; calcul des auteurs. Ce graphique présente l'évolution des plus petites corrélations de Pearson, calculées sur 3 mois en glissant, des indices DJ Stoxx sectoriels et des indices *DJ Stoxx Growth* , *Value*, *Large* et *Small*.

Nous proposons d'évaluer la dynamique des styles de gestion au moyen de classifications conditionnelles. L'idée sous-jacente au modèle est la suivante : un fonds caractérisée par un biais de gestion va surperformer le marché lorsque l'indice représentatif de ce biais domine lui-même le marché. Nous considérons donc nos 4 indices de style ainsi que l'indice général représentatif. Pour chacun des indices de style, nous retenons les périodes pour lesquelles l'indice a dominé le marché. Pour les biais *Bear* et *Bull*, nous sélectionnons le tiers des dates de la période totale pour lesquelles le marché (approximé par la moyenne transversale des rendements des fonds de la base) a été le plus et le moins performant respectivement. A

partir de ces périodes restreintes (par indice), nous effectuons à nouveau une classification en carte robuste. Puis nous indexons les classes de la carte conditionnelle en fonction de la performance des fonds représentatifs de chacune (1 pour la classe conditionnelle la moins performante, 16 pour la classe conditionnelle la plus performante). Nous reprenons ensuite notre première classification en carte robuste initiale (non conditionnelle) et déterminons, au sein de chaque classe, la proportion des individus affectés aux classes conditionnelles les plus performantes.

FIGURE 4.9. Affection de biais au moyen de classifications conditionnelles.

Biais *Bull*

0%	9%	41%	75%
7%	0%	0%	11%
23%	0%	0%	0%
87%	18%	0%	0%

Bias *Bear*

71%	45%	12%	2%
24%	33%	56%	21%
23%	0%	25%	87%
0%	3%	0%	7%

Biais *Large*

18%	36%	0%	2%
0%	0%	0%	58%
0%	0%	50%	100%
0%	0%	4%	93%

Bias *Small*

88%	18%	29%	81%
98%	11%	0%	0%
23%	0%	0%	0%
62%	0%	0%	0%

Biais *Growth*

6%	0%	0%	13%
24%	11%	0%	0%
77%	17%	6%	0%
84%	87%	61%	0%

Bias *Value*

100%	91%	94%	25%
9%	0%	22%	63%
0%	0%	6%	17%
0%	0%	0%	2%

Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006; calcul des auteurs. Sont représentés ici, la proportion des individus de chaque classe affectées *via* RSOM conditionnelle aux classes les plus performantes (14, 15, 16).

La Figure 4.9 montre les résultats obtenus par notre analyse conditionnelle lorsque l'on considère les 3 meilleurs groupes conditionnels (classe 14, 15 et 16). Les classes marquées par un biais en terme de capitalisation semblent conformes à celles précédemment déterminées. Les fonds *Large Cap* se retrouvent en bas à droite de la carte, les fonds *Small Cap* se situent en haut de la carte. Les styles *Growth* et *Value* (dont les indices représentatifs se trouvaient précédemment dans deux classes voisines) se trouvent représentés par des ensembles de classes bien distantes sur la carte.

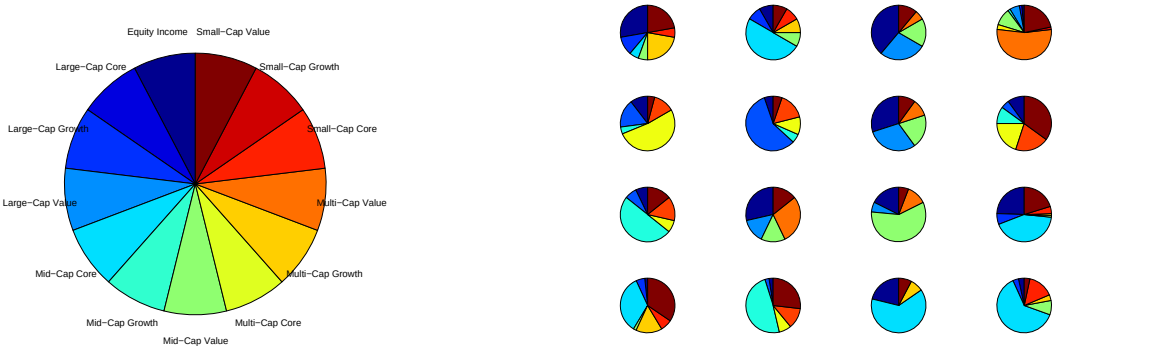
4.4.3 Recherche des biais dans l'auto-déclaration des styles des gérants

Une particularité de la base *Lipper*, est d'être fournie avec le style auto-déclaré des gérants. La Figure 4.10 présente les différentes catégories retenues. *Lipper* retient 13 styles principaux (partie gauche de la figure). Le style *Equity Income* correspond à une stratégie d'investissement dans des titres ayant un ratio dividende sur prix élevé. Selon *Lipper*, ce style s'apparente fortement au style *Large Value* ; afin de simplifier notre analyse, nous l'associons au style *Large Value*. Les 12 types de style restant sont définis en combinant les styles : *Growth*, *Value* et *Core* (absence de biais *Growth* ou *Value*) avec la taille des capitalisation : *Small Cap*, *Mid Cap*, *Large Cap* et *Multi Cap* (non contraint en terme de niveau de capitalisation). La prise en compte des scores conditionnels *Growth* et *Value* d'une part, et *Large Cap* et *Small Cap* d'autre part, nous permet d'obtenir des catégories de style comparables à celles de *Lipper*. Par exemple, pour déterminer si un fonds est *Growth*, *Value* ou *Core*, nous considérons les scores conditionnels *Growth*, *Value*. Si pour la conditionnelle *Growth*, le fonds a été affecté à l'une des trois classes les plus performantes, alors il est considéré comme tel. Nous procédons de façon simillaire pour la détermination du biais *Value*, le biais *Core* est retenu si le fonds n'est ni *Growth* ni *Value*. La recherche de biais de taille de capitalisation opère suivant le même principe. Nous confrontons alors le style auto-déclaré des gérants à celui obtenu au moyen de notre modèle.

Nous observons ainsi que 285 des 598 (soit 47%) gérants de fonds de la base communiquent un style global erroné. 118 gérants communiquent une taille de capitalisation fausse, 98 gérants communiquent un style (*Growth*, *Value* ou *Core*) inexact. Enfin 69 gérants ne respectent ni la taille de la capitalisation, ni le biais *Growth*, *Value* ou *Core* annoncés.

Les taux d'erreurs varient très sensiblement en fonction des classes considérées. Les classes 6, 7, 8, 11 et 14 sont ainsi particulièrement sujettes au biais de style proclamé. Intuitivement, l'origine de ces biais est de deux types. Il peut s'agir d'un biais structurel (en dépit du style communiqué, le gérant adopte une autre stratégie), ou d'un biais résultant des changements successifs de stratégie du gérant. L'analyse dynamique des fonds va nous

FIGURE 4.10. Superposition des auto-déclarations des styles et la classification en rendements relatifs obtenue *via* une carte robuste.



Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Sont représentées ici, les proportions des styles auto-déclarés au sein de chaque classe obtenue par classification de Kohonen.

TABLE 4.3. Taux d'erreur des styles auto-déclarés.

Classe	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Biais *Large Small*	44	62	27	35	11	33	61	80	0	6	56	35	18	74	32	25
Biais *Value* Growth	25	8	44	41	16	50	28	40	27	38	67	29	27	9	32	38

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les taux d'erreur sont exprimés en %.

permettre de différencier l'origine de ce biais.

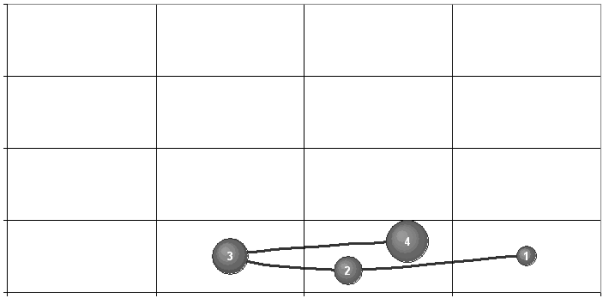
4.4.4 Stabilité de la carte des styles américains et dynamique des gérants

Dès 1996, Ferson et Schadt concluent à des changements de styles d'investissement en fonction des anticipations économiques. Il apparaît alors essentiel de prendre en compte les dynamiques des styles d'investissement (*Cf.* Chan *et alii*, 2002) lorsque l'on cherche à analyser les performances passées d'un fonds. Les modèles traditionnels du type RBSA ne permettent pas cette prise en compte. Des méthodes d'évaluation de la dynamique des styles des gérants ont déjà été proposées. Swinkels et van der Sluis (2006) utilisent, par

exemple, un filtre de Kalman pour évaluer ces changements. Même si ces modèles aboutissent à des résultats globalement satisfaisants, ils restent limités du fait des difficultés d'estimation et de calibration. Nous montrons ici, comment l'approche par les cartes de Kohonen, nous permet, de façon simple et intuitive, de suivre l'évolution des styles retenus par le gérant.

A partir de notre classification initiale, nous subdivisons notre période d'analyse en 9 semestres. Nous aurions pu choisir des sous-périodes d'analyse allant de quelques mois à un ou deux ans. Nous retenons (arbitrairement) celle du semestre dans la mesure où elle correspond assez bien à l'horizon des tactiques d'investissement des gérants. Pour chacune de ces sous-périodes, nous déterminons le positionnement des fonds sur la carte (au moyen de la projection des séries de rendements relatifs restreintes à la période considérée sur les vecteurs codes en appliquant l'équation 4.12).

FIGURE 4.11. Représentation du parcours du fonds Russell Diversified Equity Fund sur la carte.

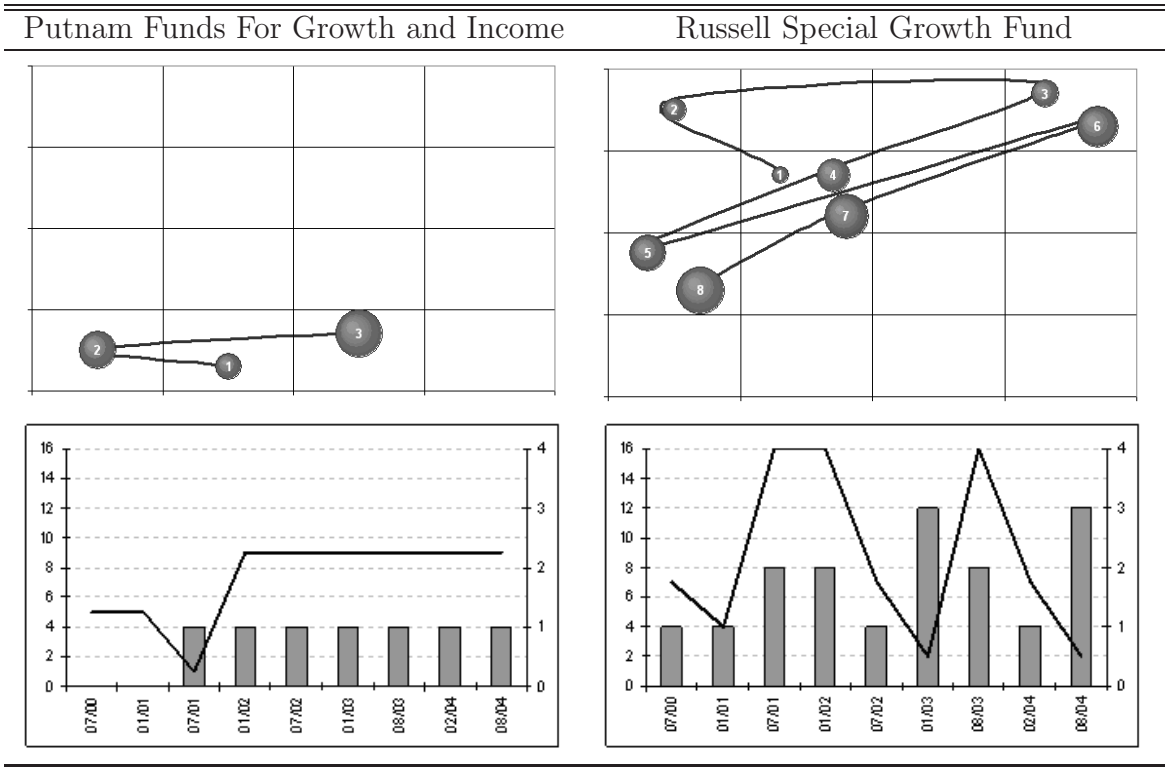


Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. La figure représente l'évolution des classifications successives du fonds Russell Diversified Equity Fund.

La Figure 4.11 illustre l'évolution du fonds *Russell Diversified Equity Fund* sur l'ensemble de la période. Le fonds est initialement classifié en classe 13 ; en mars 2003 il migre vers la classe 9, d'octobre 2005 à octobre 2006 il se positionne en classe 5 pour enfin revenir en classe 9. L'étude des distances entre les classes (*Cf.* Figure 4.6) nous a révélé que les classes 13 et 9, et 9 et 5 étaient particulièrement proches. Par ailleurs, lors de l'affectation des biais, nous avons déterminé que ces classes étaient représentatives du style *Growth*. En dépit de 3 changements de classe, nous ne pouvons pas conclure à un changement de gestion. Puisqu'un simple changement de classe ne traduit pas nécessairement un change-

ment de style nous proposons d'utiliser un critère additionnel : la distance moyenne par rapport à la classe d'origine.

FIGURE 4.12. Comparaison des parcours de deux fonds sensés appliquer un même style de gestion.



Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les graphiques du haut représentent l'évolution des classifications successives des fonds ; les graphiques du bas indiquent le positionnement dynamique du fonds (axe de gauche) et présentent l'éloignement des fonds par rapport à leurs classes d'origine (axe de droite).

Pour chaque fonds de l'univers, nous déterminons son positionnement dynamique. La Table 4.4 présente les statistiques des dynamiques des fonds de notre univers. Nous constatons que seul 4% des fonds restent dans leurs classes d'origine. Mais, comme nous l'avons précédemment évoqué, une migration d'un fonds vers une classe adjacente ne traduisant pas nécessairement un changement de style, nous considérons que la stratégie reste stable, tant que le fonds ne s'éloigne pas de plus d'une classe. A l'aune de ce critère, nous constatons que 222 gestionnaires de fonds adoptent une stratégie (relativement) constante dans le temps. La Figure 4.12 illustre les deux types de gestionnaire ; ceux adoptant une stratégie

constante (à gauche) et ceux, plus opportunistes changeant largement de style au gré des conditions de marchés.

TABLE 4.4. Dynamique des fonds, changement de classe des fonds.

Nombre de changement de classification	0	1	2	3	4	5	6	7	8
Nombre de fonds	4.0%	2.7%	11.7%	14.4%	18.1%	16.4%	14.9%	11.9%	6.0%
Durée moyenne passée dans chaque classe	9.0	4.5	3	2.2	1.8	1.5	1.2	1.1	1.0
Distance moyenne par rapport à la classe initiale	0.0	0.4	0.5	0.7	0.8	1.1	1.2	1.5	1.5

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs.

Une fois les gérants stables determinés, nous nous intéressons plus en détails à ceux changeant régulièrement de style. Nous cherchons à différencier les gérants anticipant avec succès les futures conditions de marchés (*market timer*), de ceux adoptant simplement une approche *momentum*, c'est-à-dire, suivant les tendances de marchés, les bulles spéculatives.

Nous introduisons pour cela la notion de « pari actif du gérant ». Chaque changement de gestion constaté (matérialisé par une migration sur la carte) va correspondre à un pari. Il s'agit donc de déterminer si ces paris sont gagnants ou pas. Ainsi, à chaque changement de classe de la carte, nous considérons les deux vecteurs codes représentatifs des classes (avant et après changement). Nous comparons leurs performances sur la sous-période consécutive au changement de classe. Si la performance du vecteur code de la classe de départ est inférieure à celle du vecteur code de la classe d'arrivée, alors nous considérons que le pari est gagné. Le nombre de paris gagnants rapporté au nombre de paris totaux nous permet de repérer les bon gérants. Les gérants ayant obtenu une note de 1 ont réussi tous leurs paris, ceux ayant obtenu 0 ont au contraire, manqué chacune de leurs anticipations.

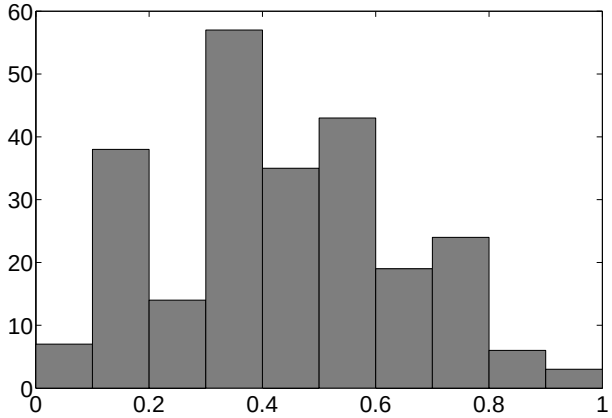
TABLE 4.5. Score dynamique des fonds.

Nombre de changement de classification	4	5	6	7	8
Score dynamique moyen	0.39	0.43	0.42	0.45	0.49

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs.

Sur l'ensemble des 376 fonds dont le style a été jugé instable, nous obtenons un score moyen de 0.44. Cette première constatation appuie notre démarche d'étude de la dynamique des styles dans la mesure où l'instabilité de ces styles induit globalement une sous performance. Néanmoins, comme l'illustre la Figure 4.13, les scores dynamiques sont très hétérogènes. Une dizaine de fonds changent de style à tort systématiquement ; alors que d'autres anticipent avec succès les futures conditions de marchés.

FIGURE 4.13. Histogramme des scores de dynamique des fonds.



Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. La figure représente la répartition des scores dynamiques des fonds.

4.5 Conclusion

Prolongeant les premiers travaux de Deboek et Kohonen (1998), et Maillet et Rousset (2003), nous avons proposé, dans cet article, une nouvelle méthodologie pour l'analyse de style des fonds. L'utilisation d'une classification robuste de Kohonen (*Cf.* Guinot *et alii*, 2006 et Maillet et Merlin, 2007) permet d'obtenir des regroupements homogènes de fonds ; chacun d'eux étant représentatif d'un style du marché. Les propriétés topologiques des cartes de Kohonen autorisent une visualisation des facteurs de risque organisés sur la carte.

Outre le caractère visuel de l'analyse, l'avantage principal de notre méthode est de

« laisser parler les données ». Nous ne dépendons pas exclusivement de la qualité informationnelle des indices de styles publiés, des catégories de classes d'actifs ou des auto-déclarations des gérants, mais nous exerçons, au contraire un certain contrôle – par comparaison – en cartographiant directement les fonds.

L'approche conditionnelle proposée, nous permet d'établir un lien entre les facteurs de risque – abstraits – obtenus et les indices de style ou sectoriels communément utilisés. Ce lien explicité, la carte de style nous permet de revérifier l'importance des biais des styles auto-déclarés des gérants. La projection dynamique des fonds sur la carte rend possible l'évaluation de la stabilité des fonds sur la carte, mais également celle de la capacité des gérants à anticiper les conditions de marché futures.

Les prochains développements de cet article seront d'appliquer notre méthode à un univers de fonds alternatifs. Les modèles d'évaluation des styles ne sont pas adaptés à cette classe d'actifs. En outre, les facteurs régissant leurs performances ne sont pas nécessairement observables (volatilité, liquidité ...). Plus spécifiquement, nous envisageons la création d'une carte de style à partir des seuls fonds alternatifs ne retenant qu'une unique stratégie. La projection dynamique des fonds dits multi-stratégies nous permettra alors d'évaluer les capacités des multi-gestionnaires à arbitrer entre les différentes stratégies alternatives.

4.6 Annexe

4.6.1 Annexe 1 : Construction des indices de style

Les indices de style sont au cœur des analyses de style communément effectuées dans les sociétés de gestion (*Return-based Style Analysis*). Il nous parait donc important de présenter les principales méthodologies de construction d'indice. Nous noterons ici que certains fournisseurs d'indices ont choisi d'être parfaitement transparents (*Standard & Poor's*, *MSCI*) quant à la méthodologie utilisée pour créer leurs indices de style ; d'autres

se montrent moins ouverts (*FTSE* et *Stoxx*).

La caractéristique *Growth* d'un titre financier est généralement mesurée au moyen de différents ratios : la croissance du chiffre d'affaire, la croissance des bénéfices prévisibles, celle des *cash flows*, l'historique de croissance des bénéfices et la croissance des actifs nets de l'entreprise. Les ratios communément utilisés pour déterminer le style *Value* sont le ratio de bénéfice sur actif net (*PER*), le cours sur actif net, le cours sur chiffre d'affaires, le cours sur *cash flows*.

Pour déterminer ses indices de style, *S&P* considère, pour un indice action donné, l'ensemble des ratios pour chacun des constituants de l'indice. Des scores *Growth* et *Value* sont déterminés pour chacun des titres. Trois variables sont utilisées pour déterminer le potentiel de croissance (le taux de croissance des bénéfices net par action, celui du chiffre d'affaire par action et le taux de croissance interne, tous trois sont mesurés sur cinq ans). Le potentiel *Value* est quant à lui déterminé à partir du cours sur actifs nets, du cours sur *cash flow*, du cours sur chiffre d'affaires et de la croissance des dividendes. Le détail des calculs permettant d'obtenir ces scores *Value* et *Growth* n'est pas communiqué. Ces scores permettent d'effectuer un classement servant à créer des indices de style pur ; les indices *Value* et *Growth* sont constitués de 33% des titres ayant obtenu les plus grands scores respectivement *Value* et *Growth*. Les 34% restants sont attribués conjointement aux deux indices, le poids attribué à chacun dépendant de l'éloignement du titre par rapport aux indices dits purs précédemment créés.

La méthodologie *MSCI* (*Cf.* MSCI, 2007) utilise elle aussi 8 variables : 3 sont utilisées pour déterminer le caractère *Value* (cours sur actifs nets, la prévision à un an des bénéfices par actions et la croissance des dividendes), et cinq pour le *Growth* (la prévision à long terme et court terme du taux de croissance du bénéfice par action, le taux de croissance interne, le bénéfices par actions historique et le taux de croissance historique du chiffre d'affaire par action). Ces différentes variables sont tout d'abord agrégées (somme pondérée)

afin d'obtenir un score *Value* et un score *Growth*. Ces deux scores sont ensuite normalisés pour en déduire quatre groupes, *Growth*, *Growth-Value*, *Value* et *Non-Growth-Non-Value*. L'ensemble des titres de l'indice global va être affecté aux indices de style en fonction de leur groupe. Les titres avec un style faiblement marqué seront affectés aux deux indices.

Concernant la méthodologie FTSE (*Cf.* FTSE, 2002), les facteurs retenus pour définir le style *Growth* sont la croissance du chiffre d'affaire et le bénéfice par action sur les trois dernières années, la projection à deux ans du chiffre d'affaire et du bénéfice par action ainsi que le *return on equity*. Les ratios utilisés pour définir le style *Value* sont les ratios de cours sur actif net, cours sur chiffre d'affaires, cours sur *cash flow* et le taux de dividende. Ces mesures sont normalisées puis agrégées (moyenne pondérée) pour chacun des titres de l'indice afin d'obtenir un score *Growth* et *Value* pour chacun d'eux. Ces mesures permettent enfin la création des indices de style (sans plus de précision à notre connaissance).

Enfin *Stoxx* utilise une approche assez différente (*Cf.* STOXX, 2007). En effet, ils considèrent uniquement 6 variables (deux historiques : cours sur actif net et taux de dividende, deux actuelles : *PER* et bénéfice par action, et deux prévus : *PER* et croissance des bénéfices) avec lesquelles ils effectuent des regroupements et en déduisent 5 groupes : les *Strong Growth*, *Weak Growth*, *Neutral*, *Weak Value* et *Strong Value*.

Tous ces indices, en raison des biais de construction et, dans une moindre mesure, des biais de sélection issus de l'indice général, vont affecter les mesures de style obtenues (*Cf.* Gallo et Lockwood, 1997).

4.6.2 Annexe 2 : Etude préliminaire, calibration de la carte et choix des inputs

Avant de développer notre modèle d'analyse de style, il nous faut déterminer le type d'input permettant l'obtention des cartes les plus exploitables. En effet, la classification obtenue va être déterminante pour l'obtention des styles des fonds d'investissement. Aussi,

nous nous proposons d'étudier la qualité de la carte en fonction de différents paramètres et traitement - par ordre d'importance : la nature des données en entrée (valeurs liquidatives, valeurs liquidatives relatives, rendements, rendements relatifs), les fonctions de distances, puis les fonctions de voisinage. Outre la taille et la forme de la carte, différents paramètres vont permettre de générer des cartes sensiblement différentes. Après énumération de ceux-ci, nous présenterons les mesures permettant d'apprécier de la qualité des différentes classes, avant d'en d'éduire les paramètres *adequats* pour la détermination des styles des fonds d'actions américaines de la base *Lipper*.

Nature des données

La nature des données classifiées va être déterminante. Comme dans toute classification, la question centrale est d'identifier les caractéristiques principales sur lesquelles établir la classification. Plusieurs choix s'offrent à nous dans le cadre de l'analyse de style de fonds. Ceux-ci peuvent en effet être caractérisé par un historique de valeurs liquidatives, de leurs rendements, ou de leurs rendements relatifs (ou encore de leurs corrélations avec les autres fonds). Dans la mesure où l'utilisation des valeurs liquidatives des fonds est sujette au problème d'échelle entre ces valeurs, le choix des variables de classification porte naturellement en premier lieu sur les rendements ou les valeurs liquidatives remise à l'échelle. Ainsi, les historiques des rendements des fonds sont utilisés tels que :

$$r_{i,t} = \frac{NAV_{i,t}}{NAV_{i,t-1}} - 1$$

pour $i = [1, ..., m]$ et $t = [2, ..., T]$, avec $NAV_{i,t}$ la valeur liquidative du fonds i, à la date t. De même, les valeurs liquidatives seront remises en base 100 au début de l'échantillon, à savoir :

$$\overline{NAV}_{i,t} = 100 \times \prod_{j=2}^{t} (r_{i,j} + 1)$$

Afin de limiter l'influence des effets de marché sur notre classification, nous choisissons également de classifier des rendements centrés par rapport à la moyenne des rendements de l'univers à chaque date. Ainsi, les rendements centrés (en coupe transversale) seront

utilisés et sont définis tels que :

$$\widetilde{r}_{i,t} = r_{i,t} - \frac{1}{n}\sum_{j=1}^{n}(r_{j,t})$$

Nous considérons également les valeurs liquidatives relatives au marché :

$$\overline{\overline{NAV}}_{i,t} = 100 \times \prod_{j=2}^{t}(\widetilde{r}_{i,j} + 1)$$

Contrairement aux paramètres précédents (fonction de voisinage et norme de proximité), le choix de la variable à classer dans le cadre d'une analyse de style est crucial comme nous le verrons, après avoir présenté dans la sous-section suivante les principaux critères de jugement de la qualité de la carte.

Choix de la fonction de distance

Le premier paramètre considéré est la fonction de distance, notée $\|.\|$ et introduite dans l'équation 4.12. Cette distance permet de déterminer le vecteur code le plus proche de l'individu sélectionné, appelé *BMU*. La mesure généralement utilisée est la distance issue de la norme Euclidienne. Toutefois, d'autres mesures peuvent améliorer sensiblement les résultats de la classification. En effet, les séries financières peuvent contenir des valeurs « anormales » sinon « aberrantes ». Il serait légitime de penser *a priori* que les valeurs particulières vont avoir un impact sensible sur les résultats de la classification. Nous considérons ainsi successivement la norme euclidienne, la norme 1, la norme infinie et enfin la mesure de corrélation. Sur la base de données étudiée, sur la période considérée et dans le cadre de l'analyse de style, aucune norme ne semble dominée les autres d'un point de vue qualitatif (meilleure lisibilité des styles) ou quantitatif (erreurs de classification plus faibles).

Choix de la fonction de voisinage

Le second paramètre permettant de modifier l'apprentissage de Kohonen est la fonction de voisinage $\Lambda(.)$ introduite dans l'équation (4.13). Quatre fonctions sont généralement envisagées dans la littérature : la fonction Gaussienne, la fonction Gaussienne tronquée, la fonction noyau d'Epanechnikov et la fonction dite à bulle :

1. la fonction Gaussienne,
$$\Lambda(x) = exp\left(-\frac{x^2}{0.1}\right) \tag{4.16}$$

2. la fonction Gaussienne tronquée,
$$\Lambda(x) = exp\left(-\frac{x^2}{0.1}\right).\mathbf{1}_{x\in[-0.25,0.25]} \tag{4.17}$$

3. la fonction noyau d'Epanechnikov,
$$\Lambda(x) = max\left(0, 1-x^2\right) \tag{4.18}$$

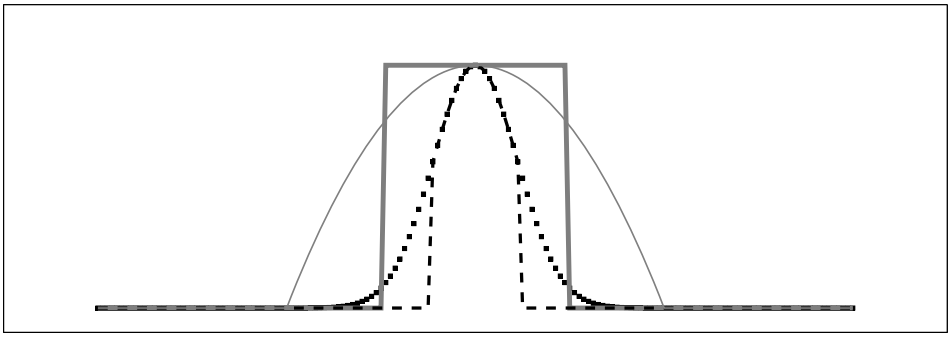
4. la fonction dite à bulle,
$$\Lambda(x) = \mathbf{1}_{x\in[-0.5,0.5]} \tag{4.19}$$

avec x la valeur issue de la fonction $neight(.,.)$ définissant le rang du voisinage entre le BMU et les neurones de la carte ($neight(BMU, \mathbf{m}_i)$ vaut 1 si le BMU et le neurons sont voisins, 2 si un neurone les sépare...)

La Figure 4.14 représente ces différentes fonctions de voisinage. Là encore, les nombreux essais menés ne sont pas concluants quant à une supériorité particulière de l'une ou l'autre des fonctions de voisinage

Critères jugeant de la qualité de la carte

Afin de juger de la qualité des différentes cartes, nous nous appuyons sur les mesures les plus largement utilisées (*Cf.* Pölzlbauer, 2004). L'erreur de quantification est généralement utilisée pour rendre compte de la qualité de regroupement. L'erreur topographique sert



Source : *données simulées* ; calcul des auteurs. La figure représente les différentes fonctions de voisinage. Les fonctions gaussienne, gaussienne tronquée, noyau d'Epanechnikov et à bulle sont représentées en pointillés noirs gras, en pointillés fins noirs, par une courbe fine grise et par une courbe épaisse grise respectivement. L'axe des abscisses a été délibérément masqué, le domaine de définition de ces fonctions évoluant au cours des itérations de l'algorithme de Kohonen.

dans le cas particulier des SOM ; elle mesure la pertinence de la projection, c'est-à-dire, la qualité de conservation topologique de la carte ainsi que son aptitude à représenter les données. Il y a clairement un arbitrage entre ces deux types de mesures. En augmentant la qualité de la projection, on dégrade celle de la topologie. D'autres mesures seraient envisageables (celles issues des normes Cosinus ou de Mahalobis) ; nous choisissons de nous limiter aux deux mesures les plus couramment utilisées, ainsi que deux mesures *ad hoc*, spécifiques à notre problème de classification de fonds, les corrélations *intra* et *extra* classe des rendements des fonds.

L'erreur de quantification (*Quantization Error*, QE) est considérée lors de l'utilisation d'algorithme de regroupement (*clustering*). Cette mesure ne tient absolument pas compte de la topologie de la carte obtenue. Il s'agit de la distance moyenne des individus avec l'élément représentatif (neurone) de leur *cluster*. Elle se formalise comme suit :

$$QE = \sum_{j=1}^{m} \left(\mathbf{x}_j - BMU_{\mathbf{x}_j} \right)^2, \tag{4.20}$$

avec m, le nombre d'individus présents dans la base d'apprentissage.

L'erreur topographique est la plus simple des mesures de préservation de la topologie.

Pour chaque individu présent dans la base de données, les deux premiers *Best Matching Units* sont déterminés. S'ils sont adjacents sur la carte, alors l'erreur associée est nulle ; sinon elle correspond au rang du voisinage des vecteurs codes considérés :

$$VE = \frac{\sum_{j=1}^{m} \left[BMU\left(\mathbf{x}_j, 1\right), BMU\left(\mathbf{x}_j, 2\right) \right]}{m}, \tag{4.21}$$

avec $V\left(\mathbf{F}_i, \mathbf{F}_j\right) = \begin{cases} 0 \text{ si } \mathbf{F}_i \text{ et } \mathbf{F}_j \text{ sont voisins sur la carte,} \\ 1 \text{ sinon,} \end{cases}$

et $BMU\left(\mathbf{x}, i\right)$ le $i^{ème}$ vecteur code le plus proche de l'individu $\mathbf{x}$.

Enfin, la mesure de distorsion est définie comme (avec les notations précédentes) :

$$d = \sum_{i=1}^{m} \sum_{j=1}^{n} \Lambda \left\{ neight \left[BMU\left(\mathbf{x}_i, j\right), \mathbf{F}_j \right] \left\| \mathbf{x}_i - \mathbf{F}_j \right\| \right\}. \tag{4.22}$$

La corrélation intra classe va nous servir à mesurer l'homogénéité de chaque groupe obtenu. Il s'agit de la corrélation moyenne des rendements des fonds appartenant à un même groupe. Pour les classes des fonds C_i pour $i = [1, ..., n]$, la corrélation *intra* se formalise comme suit :

$$c_{intra} = \frac{\sum_{j=1} \sum_{k} corr\left(\mathbf{x}_j, \mathbf{x}_k\right)}{m\left(m-1\right)}, \tag{4.23}$$

pour $j = [1, ..., m]$ tel que l'individu $\mathbf{x}_j$ appartienne à la classe C_i, $k = [1, ..., m]$ tel que l'individu $\mathbf{x}_k$ appartienne à la classe C_i et m le nombre de fonds affectés à la classe C_i.

La corrélation *extra* va nous servir à mesurer le degré de discrimination d'une classe par rapport aux autres. Nous calculons, pour mesurer cette discrimination, la moyenne des corrélations de chaque fonds d'une classe, avec les fonds hors classe. La corrélation *extra*, pour les classes des fonds C_i pour $i = [1, ..., n]$, se formalise comme suit :

$$c_{extra} = \frac{\sum_{j=1} \sum_{k} corr\left(\mathbf{x}_j, \mathbf{x}_k\right)}{m\left(t-m\right)}, \tag{4.24}$$

pour $j = [1, ..., m]$ tel que l'individu $\mathbf{x}_j$ appartienne à la classe C_i, $k = [1, ..., t-m]$ tel que l'individu $\mathbf{x}_k$ n'appartienne pas à la classe C_i et m le nombre de fonds affectés à la classe

C_i, et t le nombre d'individus classifié.

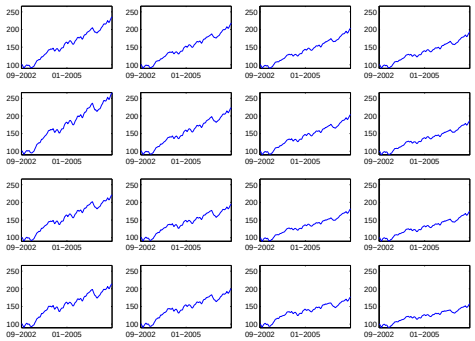
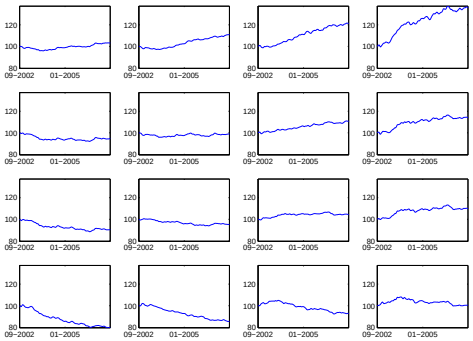
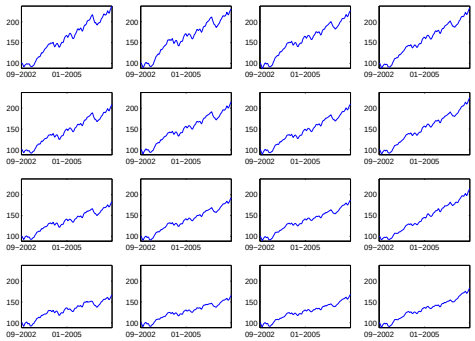
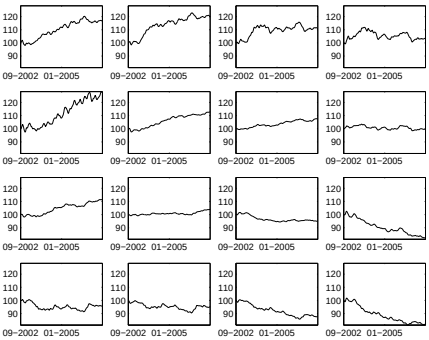
Nous avons présenté les différentes possibilités qui s'offraient à nous pour classifier un univers de fonds (types d'inputs, mesures de distance, fonction de voisinage). Les différents critères présentés vont nous permettre d'évaluer la qualité des classifications possibles et de retenir la plus appropriée pour analyser le style des fonds.

Application à la base de fonds d'actions américaines *Lipper*

La Figure 4.15 présente, pour chacune des classifications envisagées (VL, VL relatives, rendements et rendements relatifs) les évolutions des performances des fonds et des indices au sein de chaque groupe obtenu. Les caractéristiques des groupes obtenus par classification en VL et en VL relatives apparaissent assez peu diversifiées : une classe (en bas à gauche) regroupe les fonds sous performants, une autre (en haut à droite) regroupe les fonds les plus performants. Les 14 autres classes semblent assez proches. Les classes obtenues en rendements et en rendements relatifs semblent plus exploitables. Les classes sont visuellement plus discriminantes et les fonds sont mieux répartis au sein de chaque classe. La topologie des classes - l'organisation des classes - les unes par rapport aux autres semble mieux organisée pour la classification effectuée en rendements relatifs. Les fonds sous performants se situent sur la partie basse de la carte, ceux sur-performants sur la partie haute. Les fonds marqués par de fortes et faibles performances en début de période se retrouvent principalement à droite ; ceux avec une performance caractéristique en fin de période se retrouvent à gauche.

La plus intéressante des classifications dans un contexte d'analyse de style semble ainsi être obtenue en utilisant les rendements relatifs. Les différentes mesures de qualité de la carte vont nous permettre de justifier plus avant le choix de ce type de classification. Un problème se pose néanmoins concernant l'erreur de quantification ; en effet, celle-ci mesure la distance des individus (séries temporelles de rendements ou de cours de fonds), avec l'individu représentatif de leur classe. Les erreurs de quantification obtenues après classification en rendements ou en valeurs liquidatives ne seront donc pas comparables. Pour

FIGURE 4.15. Evolutions des valeurs liquidatives des individus représentatifs des seize classes suivant le type de classification utilisée.

Classification en VLs	Classification en VLs relatives
	
Classification en rendements	**Classification en rendements relatives**
	

Source : *Lipper*, base de fonds et indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Les valeurs liquidatives absolues et relatives (par rapport à la moyenne des fonds) des individus représentatifs représentées ci-dessus sont basées à 100 en début de période.

pallier le problème, nous effectuons un retraitement des données (*input* et vecteurs codes) issues des classifications en valeurs liquidatives, nous transformons ceux-ci en rendements de façon à avoir de données de natures similaires et rendre les mesures comparables.

La Table 4.6 présente les résultats obtenus. Les moyennes des corrélations *intra* classes sont du même ordre de grandeur (0.15) pour les classifications effectuées en valeurs liquidatives et en valeurs liquidatives relatives. L'utilisation des rendements pour effectuer les cartes permet d'améliorer sensiblement l'homogénéité des classes puisque nous obtenons un coefficient de corrélation *intra* classe moyen de 0.26. Enfin, les groupes de fonds classifiés

en rendements relatifs obtiennent la meilleure corrélation *intra* classe avec un coefficient moyen de 0.31. En ce qui concerne le critère de corrélation *extra* classe, il semble peu discriminant concernant le choix du type de données. En effet la corrélation *extra* est du même ordre de grandeur (proche de 0) quel que soit le type de classification. Nous notons tout de même que la classification en rendements relatifs permet l'obtention de classes marginalement plus décorrélées. Le critère de l'erreur de quantification (modifiée comme décrit ci-dessus) est également assez peu discriminant ; quel que soit le type de classification adoptée, celle-ci vaut approximativement 0.04. Là encore, l'erreur est légèrement inférieure pour la classification en rendements relatifs. Le critère d'erreur topographique permettant de juger de la cohérence et de la stabilité de la carte en terme de voisinage, présente la classification en rendements relatifs comme la plus cohérente. En dépit du fait que ces critères mesurent et quantifient des propriétés tout à fait différentes, ils nous conduisent à adopter une classification en rendements relatifs.

TABLE 4.6. Mesures d'appréciation des cartes en fonction du type de données choisi exprimées en pourcentages

Mesures	Valeurs liquidatives	Valeurs liquidatives relatives	Rendements	Rendements relatifs
Moyenne des corrélations *intra* classe	15.52	15.71	25.78	30.59
Moyenne des corrélations *extra* classe	10.73	-0.58	-2.19	-2.46
Erreur de quantification	4.62	4.60	4.49	4.45
Erreur topographique	20.08	15.35	4.29	2.27

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Cette table présente les mesures d'appréciation des cartes en fonction du type d'inputs.

La Table 4.7 présente les erreurs topographiques et de quantification obtenues suivant les différentes fonctions retenues, qui sont du même ordre de grandeur quelle que soit la classification considérée. Les différences n'étant pas suffisamment significatives, nous choisissons de conserver la calibration la plus courante de la carte : une fonction de voisinage gaussienne, avec la norme euclidienne comme fonction de distance.

TABLE 4.7. Mesure de la qualité des classifications obtenues en rendements relatifs (exprimées en pourcentages)

Fonction de voisinage	Fonction de distance	Erreur topographique	Erreur de quantification	Mesure de distortion
Bulle	Norme Euclidienne	7.30	1.20	3.48
	Norme infinie	7.55	0.30	3.45
	Norme 1	7.16	0.30	3.25
	Corrélation	7.35	0.90	3.36
Gaussienne	Norme Euclidienne	7.24	1.20	3.35
	Norme infinie	7.57	2.11	3.22
	Norme 1	7.19	2.41	3.26
	Corrélation	7.43	1.51	3.49
Gaussienne tronquée	Norme Euclidienne	7.21	1.81	3.31
	Norme infinie	7.54	5.42	3.24
	Norme 1	7.32	9.94	3.42
	Corrélation	7.40	2.41	3.47
Epanechnikov	Norme Euclidienne	7.20	0.30	3.29
	Norme infinie	7.49	0.60	3.38
	Norme 1	7.24	2.71	3.26
	Corrélation	7.41	1.81	3.38

Source : *Lipper*, base d'indices américains en dollar, données hebdomadaires (valeur vendredi) du 30/08/2002 au 29/12/2006 ; calcul des auteurs. Cette table présente les mesures d'appréciation des cartes suivant les fonctions de voisinage et norme choisies.

Conclusion Générale

Les quatre essais présentés dans cette thèse ambitionnent de répondre à certaines problématiques auxquelles praticiens et chercheurs académiques sont couramment confrontés lors de l'application de modèles financiers à des classes d'actifs comprenant des fonds alternatifs : les prétraitements des bases, la sélection de portefeuilles optimaux, la détection de valeurs aberrantes, la caractérisation des styles des fonds alternatifs.

Les spécificités des fonds alternatifs rendent en effet nécessaire la considération d'un cadre d'analyse qui leur soit adapté. La taille et la liquidité des actifs investis, la transparence, le système de rémunération et les stratégies employées par les gérants, le cadre légal et les clauses contractuelles différencient largement cette classe d'actifs des fonds d'investissements traditionnels. Les particularités des *hedge funds* induisent en particulier que l'hypothèse de normalité des rendements ne peut être retenue. Ce rejet du postulat Gaussien remet alors en cause bon nombre de modèles de la finance moderne (*Cf.* Hocquard *et alii*, 2008).

Dans le premier chapitre, nous avons traité du problème de non-renseignement de certaines valeurs dans les bases de données financières (*Cf.* Malhotra, 1987 ; diCesare, 2006). De nombreux modèles requièrent en effet de disposer d'un ensemble d'observations complet. La nécessité d'un traitement préalable des données est accrue lorsque l'on considère des fonds alternatifs. Le cadre légal peu contraignant dont ils disposent et le fait qu'ils ne soient pas soumis à l'obligation de communiquer (toutes) leurs performances rendent les bases de données de fonds alternatifs particulièrement sujettes à ce problème de séries

incomplètes. Ainsi, la première application qu'il nous a semblé utile de présenter propose d'estimer et de remplacer ces valeurs manquantes.

Les modèles de complétion existants (à notre connaissance) sont soit trop destructeurs d'information, soit inappropriés en raison des hypothèses qu'ils sous-tendent. Les cartes de Kohonen, combinées aux fonctions orthogonales empiriques et à l'algorithme de génération aléatoire contrainte rendent possible la réalisation d'un modèle de complétion de données respectant les caractéristiques des rendements des fonds alternatifs. Conjecturant qu'une réalisation future ne constitue – finalement - qu'une valeur manquante, nous avons ensuite montré comment les cartes de Kohonen pouvaient également être utilisées pour développer un modèle de génération de *scénarii*.

Nous avons appliqué nos modèles de complétion et de simulation à un ensemble de fonds alternatifs. Nous nous sommes attachés à illustrer les conséquences de l'utilisation de modèles inadaptés (l'algorithme « *Expectation-Conditional-Maximisation* » et ceux de simulation de Monte Carlo en particulier) sur des modèles usuels en finance d'allocation d'actifs et de mesures de risque. Différents critères nous ont permis d'évaluer l'aptitude de nos modèles à générer des données qui respectent les caractéristiques des séries. Nous avons tout d'abord considéré l'erreur quadratique des moments associés aux différentes séries reconstruites. Par rapport aux modèles standard, nous avons pu réduire très sensiblement l'impact des biais de reconstruction et de simulation (amélioration de 80%) sur les quatre premiers moments. Nous avons ensuite montré les conséquences de ces erreurs sur la composition de portefeuilles de référence (ratio de Sharpe *maximum* et de volatilité *minimale* ; *Cf.* Zhou, 2007 ; Siegel et Woodgate, 2007). L'impact de ces biais sur plusieurs mesures de risque a également été étudié. Nous établissons qu'à l'exception des *VaR* Gaussiennes, les mesures obtenues sont moins biaisées après l'application de nos méthodes. Nous avons ainsi montré l'importance de retenir une méthodologie dont les hypothèses s'accordent avec les données exploitées, et la nécessité d'écarter certains modèles standard lorsque des fonds alternatifs sont considérés.

A ce stade, les résultats obtenus nous incitent à poursuivre nos travaux dans cet axe de recherche. S'inspirant des travaux de Figueroa García *et alii* (2008), les propriétés d'auto-

corrélation des séries de rendements de fonds alternatifs pourront notamment être intégrées à l'algorithme de génération aléatoire contrainte. Quant à la méthode de simulations, des tests sur la fiabilité des mesures de risques obtenues pourront être réalisés (*Cf.* Filipović et Vogelpoth, 2007). Il sera également intéressant d'évaluer la qualité de nos simulations au regard de techniques rejetant la normalité des rendements des actifs (*Cf.* Hörmann *et alii*, 2008 ; Behr et Pötter, 2009). Notre approche validée, il sera possible en adoptant un cadre conditionnel de considérer uniquement les *scenarii* de type *stress tests*, et d'en déduire des mesures de risques extrêmes. Enfin, si les *scenarii* générés respectent fidèlement les dépendances, nous les combinerons aux méthodes de détermination de la copule adéquate (*Cf.* Durrleman *et alii*, 2000). En effet, les données simulées permettront alors d'améliorer significativement l'inférence et la puissance des tests de validation du choix de la bonne spécification des copules (en anglais, *Goodness-of-Fit* ; *Cf.* Fermanian et Scaillet, 2002 ; Genest *et alii*, 2009).

Dans le deuxième chapitre, nous nous sommes intéressés aux modèles de sélection et d'optimisation de portefeuille. Dans le cadre d'analyse élargi à un critère espérance-variance-asymétrie-*kurtosis*, l'approche duale est généralement retenue. En effet, les fonctions d'utilité quartique autorisent une prise en compte simple des préférences pour les moments d'ordre supérieur (*Cf.* Christodoulakis et Peel, 2006). L'approche primale est par contre largement complexifiée lorsque l'on considère les préférences des agents pour l'asymétrie positive et une certaine défiance vis-à-vis des risques extrêmes (*Cf.* Maringer et Parpas, 2009). En effet, dans le plan espérance-variance, l'univers des portefeuilles constitue un ensemble convexe, il est ainsi possible de définir une relation d'ordre pour obtenir l'ensemble des portefeuilles efficients. Dès lors que nous considérons les moments d'ordre supérieur, cet ensemble n'est pas nécessairement convexe.
L'introduction de la fonction de pénurie nous a permis de réunifier les deux approches primale et duale (*Cf.* Briec *et alii*, 2007). Nous avons notamment pu expliciter le lien existant entre la direction de projection de la fonction de pénurie retenue et les préférences des agents. Par ailleurs, pour limiter les problèmes d'estimation des moments des porte-

feuilles, nous avons retenu un ensemble de statistiques robustes - les *L-moments*, qui se substituent favorablement aux moments traditionnels. Nous avons également pu reformuler le problème d'optimisation de portefeuille à partir de vecteurs de *co-L-moments*. Cette réécriture nous a notamment permis de réduire les temps d'exécution des optimisations de portefeuille de manière significative.

Considérant une base d'actions européennes, nous avons obtenu un ensemble de portefeuilles optimaux correspondant aux préférences quartiques d'agents rationnels. Cet ensemble nous a conduits à représenter les surfaces efficientes multidimensionnelles, et à analyser les interactions entre les différents moments. Ainsi des arbitrages de l'asymétrie par rapport aux deux premiers moments semblent réalisables, la *kurtosis* fortement liée à la volatilité semble par contre négligeable lors de la recherche de portefeuilles optimaux. Pour des niveaux d'asymétrie fixés, nous avons retrouvé les formes convexes des frontières efficientes du cadre traditionnel moyenne-variance. La considération de fonctions d'utilité générales nous a ensuite amené à l'étude des préférences d'agents rationnels (*Cf.* Mitton et Vorkink, 2007 ; Crainich et Eeckhoudt, 2008). La prise en compte des moments d'ordre supérieur ne semble modifier que très marginalement le choix des agents rationnels. L'intensification des préférences des agents pour des asymétries positives et de faibles *kurtosis* nous a néanmoins permis de vérifier le caractère négligeable du quatrième moment.

La première amélioration envisagée sera de formuler les (co-)moments d'ordre supérieur de sorte que les termes redondants soient supprimés. Nous serons alors en mesure de réduire à nouveau la taille des matrices de co-moments, et donc le temps d'exécution nécessaire à la détermination des moments des portefeuilles (*Cf.* Jondeau *et alii*, 2007). Il sera également intéressant d'appliquer notre méthodologie pour résoudre des problèmes d'allocation de fonds de fonds alternatifs (*Cf.* Bergh et van Rensburg, 2008). Enfin, une prise en compte des dynamiques des co-moments d'ordre supérieur est également prévue (*Cf.* Guidolin et Timmermann, 2008) dans le cadre de la définition de stratégies basées sur les prévisions des densités conditionnelles de rentabilités (« *distribution timing* », *Cf.* Jondeau et Rockinger, 2008).

Le chapitre 3 est consacré à l'étude de l'impact des données aberrantes sur les modèles d'allocation d'actifs (*Cf.* Michaud et Michaud, 2008). Les séries temporelles d'actifs financiers sont en effet sujettes à de nombreuses erreurs. Dans la mesure où les moments d'ordre supérieur sont particulièrement sensibles à ces erreurs (*Cf.* Serfling, 2008), il était naturel de s'attendre à ce que les modèles d'allocation d'actifs intégrant explicitement l'asymétrie et la *kurtosis* des portefeuilles soient particulièrement sensibles à ces valeurs. Nous avons illustré la nécessité d'appliquer un prétraitement pour la correction des valeurs aberrantes avant d'utiliser ces données dans des modèles d'allocation d'actifs.

Nous avons présenté un modèle de détection d'*outlier* défini à partir d'une modélisation de la volatilité (*Cf.* Franses et Ghijsels, 1999 ; et Charles et Darné, 2005) par un processus de type *ANN-GARCH* (*Cf.* Donaldson et Kamstra, 1997 ; Miazhynskaia *et alii*, 2006 ; et Bildirici et Ersin, 2009). Le perceptron multi-couches rend possible l'explication et la prévision d'une partie du résidu non prise en compte par un modèle *GARCH*. Ainsi, l'introduction du *ANN-GARCH* nous a permis d'améliorer sensiblement le modèle de détection et de correction des valeurs aberrantes proposé par Charles et Darné (2005).

Afin d'illustrer et de tester notre modèle, nous nous sommes attachés à corriger les valeurs aberrantes présentes dans une base de données constituée des cours quotidiens des principales actions incluses dans l'indice CAC40. La période retenue (du 1er janvier 1996 au 21 janvier 2009) comporte de nombreux évènements extrêmes de marché. Cette base nous a conduits à déterminer plus de 184 rendements aberrants. Nous avons alors cherché à déterminer l'origine de ces rendements extrêmes. Une première démarche qualitative (relative à une recherche de dépêches concernant les entreprises) nous a permis de vérifier que la plupart correspondait à la réalisation d'un risque spécifique. Nous avons complété cette première analyse par une démarche quantitative. La comparaison des indicateurs de turbulence de marché calculés sur l'indice du CAC 40 et sur chacun des titres nous a permis de vérifier cette première conclusion.

Nous avons ensuite proposé une étude comparative des portefeuilles efficients obtenus en appliquant notre méthode de sélection de portefeuilles fondée sur les fonctions de pénurie en considérant des moments traditionnels ou robustes (*L-moments*) lorsque les données

sont brutes ou corrigées par notre algorithme de détection des données aberrantes. Retenant quatre portefeuilles caractéristiques de la frontière efficiente étendue, nous avons procédé à des simulations dynamiques d'allocation d'actifs. Nous avons finalement évalué l'impact des rendements aberrants sur les modèles d'allocation d'actifs au regard des caractéristiques *ex ante* des portefeuilles obtenus. Il apparaît ainsi qu'à l'exception du portefeuille de volatilité minimum, les portefeuilles efficients sont particulièrement sensibles aux *outliers*. Le retraitement des rendements par la méthode des *ANN-GARCH* apparaît plus efficace que la considération des statistiques robustes. Néanmoins, la combinaison de ces deux approches nous a permis d'améliorer significativement les caractéristiques ex post des portefeuilles simulés.

Nous envisageons deux pistes majeures d'améliorations. La prise en compte explicite de la dynamique des moments d'ordre supérieur devrait permettre une amélioration significative de nos résultats. De façon analogue au chapitre 2, la définition de stratégies basées sur les prévisions des densités conditionnelles de rentabilités (« *Distributional Timing* », *Cf.* Jondeau et Rockinger, 2008) nous permettra d'illustrer encore plus franchement l'impact des rendements aberrants. La composante neuronale de notre modèle pourra également intégrer la dynamique des moments en considérant un perceptron multi-couches dit récurrent (*Cf.* Miazhynskaia *et alii*, 2006).

Dans le chapitre 4, nous avons finalement proposé un modèle d'analyse de style qui est potentiellement applicable aux fonds alternatifs (*Cf.* Fung et Hsieh, 2002 ; Kuenzi et Shi, 2007). L'opacité caractéristique de la gestion alternative renforce le besoin d'explication et d'analyse des stratégies retenues par les gérants (*Cf.* Monarcha, 2008). Une approche statistique fondée sur l'analyse des performances passées a été retenue pour évaluer le style des fonds. Nous avons présenté les principaux modèles de style, puis nous nous sommes attachés à en rappeler les limites. L'hypothèse la plus restrictive (selon nous) réside dans le fait que les analyses standard sont statiques ; les sensibilités des fonds aux différents facteurs de style sont supposées constantes dans le temps (*Cf.* Annaert et van Campenhout, 2007). Cette dernière hypothèse apparaît fragile dans la mesure où non seulement le gérant

peut être amené à changer de tactique (*Cf.* Basak *et alii*, 2007 ; Ainsworth *et alii*, 2008), mais les titres eux-mêmes sont sujets à des variations de style (*Cf.* Fama et French, 2007).

Nous avons ainsi retenu les cartes de Kohonen pour fonder notre modèle d'analyse de style (*Cf.* Maillet et Rousset, 2003 ; Aaron *et alii* 2004 et 2005). A partir d'une classification robuste de fonds, nous avons pu extraire directement les facteurs de style de la base de fonds considérée. Les problèmes de variations du style des titres ainsi que ceux liés à l'identification des facteurs probants ont ainsi pu être contourné. Adoptant une approche conditionnelle, nous avons pu caractériser les facteurs obtenus en termes d'actifs réels de l'économie. Nous nous sommes ensuite intéressés à la prise en compte de la dynamique des styles retenus par les gérants (*Cf.* Basak *et alii*, 2007 ; Baghai-Wadji *et alii*, 2008 ; Huang *et alii*, 2008). Les projections successives des fonds sur les facteurs générés par la carte nous a permis d'évaluer cette dynamique. Au travers de cette dernière analyse, nous avons ainsi pu apprécier la stabilité des types de gestion adoptés par les gérants.
Afin de valider notre approche, nous avons considéré un univers de fonds investis en actions américaines pour lesquels nous disposions d'informations qualitatives, en particulier le style auto-déclaré des gérants. Nous avons ainsi constaté qu'un tiers des stratégies annoncées ne correspondaient pas exactement à celles obtenues. L'analyse dynamique des styles des fonds nous a également permis d'identifier les gestionnaires démontrant une certaine aptitude à anticiper les orientations de marché et les styles dominants (*Cf.* Cremers et Petajisto, 2009 ; Huang *et alii*, 2008).

Cette analyse réalisée sur un ensemble de fonds raisonnablement transparents constitue une première étape de notre travail de recherche sur les styles d'investissement. Nous prévoyons d'appliquer une analyse similaire sur un univers de fonds alternatifs. Nous porterons alors une attention particulière aux évolutions des couvertures et des leviers employés par les gestionnaires (*Cf.* Bodson *et alii*, 2008). Plus généralement, l'analyse des dérives de gestion (*style drifts* en anglais ; *Cf.* Ainsworth *et alii*, 2008) nous conduira à revisiter les modèles d'attribution de performance des fonds alternatifs. A travers l'étude dynamique des stratégies des gérants, nous tâcherons de détecter des engouements pour un style spécifique. Nous pourrions alors compléter les récents travaux réalisés sur l'origine

et l'impact des bulles spéculatives (*Cf.* Bird *et alii*, 2008 ; Dorn *et alii*, 2008). Une autre extension de notre modèle consistera à inclure des informations qualitatives à notre base de classification. En effet, les cartes de Kohonen autorisent la classification de données hétérogènes (qualitatives et quantitatives). Nous disposerons alors d'un cadre unifié pour intégrer les approches traditionnelles de type *RBSA* et celles plus récentes (*Cf.* Brown *et alii*, 2009) consistant à évaluer les performances et le risque de *hedge funds* au regard d'informations résultant des audits opérationnels et juridiques (*due diligences* en anglais).

Références bibliographiques

Aaron C., I. Bilon, S. Galanti et Y. Tadjeddine, (2005), "Les styles de gestion de portefeuille existent-ils ?", *Revue d'Economie Financière 81*, 171-188.

Aaron C., S. Galanti et Y. Tadjeddine, (2004), "La gestion collective dans un march agité : la dynamique des styles de gestion à partir des cartes de Kohonen", *Revue d'Economie Politique 114*, 507-526.

Abreu D. et M. Brunnermeier, (2003), "Bubbles and Crashes", *Econometric Society 71*, 173-204.

Ackerman C., R. McEnally et D. Ravenscraft, (1999), "The Performance of Hedge Funds: Risk, Return and Incentives", *Journal of Finance 54*, 833-874.

Adcock C., (2004), "Capital Asset Pricing for UK Stocks under the Multivariate Skew-Normal Distribution", dans *Skew Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton Eds, Chapman & Hall, 191-204.

Adcock C., (2008), "Asset Pricing and Portfolio Selection based on the Multivariate Skew-Student Distribution", Actes du colloque *FFM08*, 30 pages.

Agarwal V. et N. Naik, (2000), "Multi-period Performance Persistence Analysis of Hedge Funds", *Journal of Financial and Quantitative Analysis 35*, 327-342.

Agarwal V. et N. Naik, (2004), "Risks and Portfolio Decisions Involving Hedge Funds", *Review of Financial Studies 17*, 63-98.

Agren M., (2006), "Prospect Theory and Higher Moments", *Working Paper, Université d'Uppsala*, 29 pages.

Ainsworth A., K. Fong et D. Gallagher, (2008), "Style Drift and Portfolio Management for Active Australian Equity Funds", *Australian Journal of Management 32*, 387-418.

Alexiev J., (2005), "The Impact of Higher Moments on Hedge Fund Risk Exposure", *Journal of Alternative Investments 7*, 50-65.

Alizadeh S., M. Brandt et F. Diebold, (2002), "Range-based Estimation of Stochastic Volatility Models", *Journal of Finance 57*, 1047-1091.

Allison P., (2001), *Missing data. Thousand Oaks*, Sage Publications, 93 pages.

Amin G. et H. Kat, (2003-a), "Stocks, Bonds and Hedge Funds", *Journal of Portfolio Management 29*, 113-120.

Amin G. et H. Kat, (2003-b), "Hedge Fund Performance: 1990-2000: Do the Money Machines Really add Value?", *Journal of Financial and Quantitative Analysis 38*, 1-23.

Ammann M. er M. Verhofen, (2006), "The Effect of Market Regimes on Style Allocation", *Financial Markets and Portfolio Management 20*, 309-337.

Andersson J., (2001), "On the Normal Inverse Gaussian Stochastic Volatility Model", *Journal of Business & Economic Statitics 7*, 297-305.

Ané T., L. Ureche-Rangau, J.-B. Gamert et J. Bouverot, (2008), "Robust Outlier Detection for Asia-Pacific Stock Index Returns", *Journal of International Financial Markets, Institutions and Money 18*, 326-343.

Annaert J. et G. van Campenhout, (2007), "Time Variation in Mutual Fund Style Exposure", *Review of Finance 11*, 633-661.

Arditti F., (1967), "Risk and the Required Return on Equity", *Journal of Finance 22*, 19-36.

Arnold B. et G. Meelden, (1975), "Characterization of Distributions by Sets of Moments of Order Statistics", *Annals of Statistics 3*, 754-758.

Arrow K., (1964), "The Role of Securities in the Optimal Allocation of Risk-bearing", *Review of Economic Studies 31*, 91-96.

Arrow K., (1970), *Esays in the Theory of Risk-bearing*, North-Holland Publications, 278 pages.

Artzner P., F. Delbaen, J. Eber et D. Heath, (1999), "Coherent Measures of Risk", *Mathematical Finance 9*, 203-228.

Asness C., R. Krail et J. Liew, (2001), "Do Hedge Funds Hedge?", *Journal of Portfolio Management 28*, 6-19.

Athayde (de) G. et R. Flôres, (1999), "Introducing Higher Moments in the CAPM: Some Basic Ideas", *EPGE Working Paper 362*, 16 pages.

Athayde (de) G. et R. Flôres, (2002), "Portfolio Frontier with Higher Moments: The Undiscovered Country", *Computing in Economics and Finance 209*, 1-42.

Athayde (de) G. et R. Flôres, (2003), "Incorporating Skewness and *Kurtosis* in Portfolio Optimization: A Multi-dimensional Efficient Set", dans *Advances in Portfolio Construction and Implementation*, Satchell-Scowcroft Eds, Butteworth-Heinemann, 243-257.

Athayde (de) G. et R. Flôres, (2004), "Finding a Maximum Skewness Portfolio a General Solution to Three-moments Portfolio Choice" , *Journal of Economic Dynamics and Control 28*, 1335-1352.

Athayde (de) G. et R. Flôres, (2006), "On Certain Geometric Aspects of Portfolio Optimisation with Higher Moments", dans *Multi-moment Asset Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 37-50.

Avouyi-Dovi S. et R. Caulet, (1995), "Les réseaux de neurones artificiels : une application à la prévision des prix des actifs financiers", *Caisse des Dépôts et Consignations, Cahiers de Recherche 1995-19/T*, 23 pages.

Baghai-Wadji R., R. El-Berry, S. Klocker et M. Schwaiger, (2008), "Changing Investment Styles: Style Creep and Style Gaming in the Hedge Fund Industry", *International Journal of Intelligent Systems in Accounting and Finance Management 14*, 157-177.

Baillie R. et T. Bollerslev, (1989), "The Message in Daily Exchange Rates: A Conditional-Variance Tale", *Journal of Business & Economic Statistics 7*, 297-305.

Baker R., (2008), "An Order-statistics-based Method for Constructing Multivariate Distributions with Fixed Marginals", *Journal of Multivariate Analysis 99*, 2312-2327.

Balke N. et T. Fomby, (1994), "Large Shocks, Small Shocks, and Economic Fluctuations: Outliers in Macroeconomic Time Series", *Journal of Applied Econometrics 9*, 181-200.

Banz R., (1981), "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics 9*, 3-18.

Barberis N., (2000), "Investing in the Long-run when Returns are Predictable", *Journal of Finance 55*, 225-264.

Barberis N. et M. Huang, (2007), "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices", *Working Paper, National Bureau of Economic Research*, 42 pages.

Barberis N. et A. Shleifer, (2003), "Style Investing", *Journal of Financial Economics 68*, 161-199.

Barndorff-Nielsen O., (1997), "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling", *Scandinavian Journal of Statistics 24*, 1-13.

Barnett V. et T. Lewis, (1978), *Outliers in Statistical Data*, Wiley, 604 pages.

Barone L., (2008), "Bruno de Finetti and the Case of the Critical Line's Last Segment", *Insurance: Mathematics and Economics 42*, 359-377.

Basak S., A. Pavlova et A. Shapiro, (2007), "Optimal Asset Allocation and Risk Shifting in Money Management", *Review of Financial Studies 20*, 1583-1621.

Basu S., (1983), "The Relationship between Earning Yield, Market Value and the Return for NYSE Common Stocks", *Journal of Financial Economics 12*, 126-156.

Bayazit M. et B. Önöz, (2002), "LL-moments for Estimating Low Flow Quantiles", *Hydrological Sciences 47*, 707-720.

Beedles W. et M. Simkowitz, (1978), "Diversification in a Three-moment World", *Journal of Financial and Quantitative Analysis 13*, 927-941.

Behr A. et U. Pötter, (2009), "Alternatives to the Normal Model of Stock Returns: Gaussian Mixture, Generalised Logf and Generalised Hyperbolic Models", *Annals of Finance 5*, 49-68.

Ben Dor A., R. Jagannathan et I. Meier, (2003), "Understanding Mutual Funds and Hedge Funds Styles Using Return-Based Style Analysis", *Journal of Investment Management 1*, 97-137.

Ben Omrane W. et E. de Bodt, (2007), "Using Self-Organizing Maps to adjust for Intra-day Seasonality", *Journal of Banking and Finance 31*, 1817-1838.

Ben-Zvi A. et B. Azmon, (1997), "Joint Use of L-moment Diagram and Goodness-of-Fit Test: A Case Study of Diverse Series", *Journal of Hydrology 198*, 245-259.

Benishay H., (1987), "A Fourth-degree Polynomial Utility Function and its Implications for Investors Responses Toward Fourth Moments of the Wealth Distribution", *Journal of Accounting, Auditing and Finance 2*, 203-238.

Benishay H., (1989), "More Weigths to the Friedman-Savage Hypothesis: a Comment", *Journal of Accounting, Auditing and Finance 4*, 518-527.

Benishay H., (1992), "The Pratt-Arrow Requirement in a Fourth Degree Polynomial Utility Function", *Journal of Accounting, Auditing and Finance 7*, 97-115.

Benson K., R. Faff et P. Pope, (2003), "The Relevance of Investor Risk Classes in Ranking: Fund Performance: An Application of the Extended Mean-Gini CAPM", *Journal of Quantitative Economics 1*, 20-35.

Berényi Z., (2001), "Performance of Leveraged Asset Funds", *Working Paper, Université de Munich*, 42 pages.

Berényi Z., (2002), "Measuring Hedge Funds' Risks with Moment-based Variance-equivalent Measures", *Working Paper, Université de Munich*, 35 pages.

Bergh G. et P. van Rensburg, (2008), "Hedge Funds and Higher Moment Portfolio Selection", *Journal of Derivatives & Hedge Funds 14*, 102-126.

Bernoulli D., (1738), traduit en anglais par L. Sommer, (1954), "Exposition of a New Theory on the Measurement of Risk", *Econometrica 22*, 23-36.

Bernoulli D., (1777), traduit en anglais par G. Allen, (1961), "The most Probable Choice between Several Discrepant Observations and the Formation therefrom of the most Likely Induction", *Biometrika 48*, 1-18.

Best M. et R. Grauer, (1992), "Sensitivity Analysis for Mean-Variance Portfolio Problems", *Management Science 37*, 980-989.

Bildirici M. et Ö. Ersin, (2009), "Improving Forecasts of GARCH Family Models with the Artificial Neural Networks: An Application to the Daily Returns in Istanbul Stock Exchange", *Expert Systems with Applications 36*, 7355-7362.

Bird R., L. Casavecchia et P. Woolley, (2008), "Insights into the Market Impact of Different Investment Styles", *Working Paper, Université Technologique de Sydney*, 24 pages.

Bishop C., (1995), *Neural Networks for Pattern Recognition*, Clarendon Press, 504 pages.

Black F., (1976), "The Pricing of Commodity Contracts", *Journal of Financial Economics 3*, 167-179.

Black F. et M. Scholes, (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy 81*, 637-654.

Blanchard O. et M. Watson, (1982), "Bubbles, Rational Expectations and Financial Markets", paru dans *Crises in the Economic and Financial Structure*, Watchtel Eds, Lexington Books, 295-316.

Bodson L., A. Coën et G. Hübner, (2008), "Dynamic Hedge Fund Style Analysis with Errors-in-Variables", Actes du colloque *Multinational Finance Society*, 29 pages.

Bollerslev T., (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics 31*, 307-327.

Bollerslev T., (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Returns", *Review of Economics and Statistics 69*, 542-547.

Bollerslev T., R. Engle et J. Wooldridge, (1988), "A Capital Asset Pricing Model with Time-Varying Covariances", *Journal of Political Economy 96*, 116-131.

Bollerslev T., R. Chou et K. Kroner, (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics 52*, 5-59.

Bonhomme S. et J.-M. Robin, (2008), "Consistent Noisy Independent Component Analysis", *Working Paper, Université de Paris-1*, 45 pages.

Borland L., J.-P. Bouchaud, J.-F. Muzy et G. Zumbach, (2005), "The Dynamics of Financial Markets – Mandelbrot's Multifractal Cascades, and Beyond", *Science & Finance Working Paper 500061*, 24 pages.

Bouchaud J., P. Cizeau et L. Laloux, (1999), "Noise Dressing of Financial Correlation Matrices", *Physical Review Letters 83*, 1467-1470.

Bowman A. et A. Azzalini, (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-plus Illustrations*, Oxford Science Publications, 208 pages.

Boyd J., E. Kennelly et P. Pistek, (1994), "Estimation of EOF Expansion Coefficients from Incomplete Data", *Deep Sea Research 41*, 1479-1488.

Boyle P., (1977), "Options: A Monte Carlo Approach", *Journal of Financial Economics 4*, 323-338.

Brand M., (2002), "Incremental Singular Value Decomposition of Uncertain Data with Missing Values", Actes du colloque *ECCV 2350*, 707-720.

van den Branden K. et S. Verboven, (2009), "Robust Data Imputation", *Computational Biology and Chemistry 33*, 7-13.

Brennan M. (1993), "Agency and Asset Pricing", *Working Paper, Université de Californie*, 35 pages.

Briec W. et K. Kerstens, (2005), "Markowitz Portfolio Selection in Multidimensional Moment Space", *Working Paper, Université de Perpignan*, 28 pages.

Briec W. et K. Kerstens, (2007), "Portfolio Selection in Mutlidimensional General and Partial Moment Space", *LEM Working Paper, Université de Lille*, 32 pages.

Briec W. et K. Kerstens, (2009), "Multi-horizon Markowitz Portfolio Performance Appraisals: A General Approach", *Omega 37*, 50-62.

Briec W., K. Kerstens et O. Jokung, (2007), "Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach", *Management Science 53*, 135-149.

Briec W., K. Kerstens et J. Lesourd, (2004), "Single-Period Markowitz Portfolio Selection, Performance Gauging, and Duality: A Variation on the Luenberger Shortage Function", *Journal of Optimization Theory and Applications 120*, 1-27.

Bris A., W. Goetzmann et N. Zhu, (2007),"Efficiency and the Bear: Short Sales and Markets around the World", *Journal of Finance 62*, 1029-1079.

Britten-Jones M., (1999), "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights", *Journal of Finance 54*, 655-671.

Brockett P. et J. Garven, (1998), "A Reexamination of the Relationship between Preferences and Moment Orderings by Rational Risk Averse Investors", *Geneva Papers on Risk and Insurance Theory 29*, 127-137.

Brockett P. et J. Golden, (1987), "A Class of Utility Functions containing all the Common Utility Functions", *Management Science 33*, 955-964.

Brockett P. et Y. Kahane, (1992), "Risk, Return, Skewness and Preference", *Management Science 38*, 851-866.

Brooks C. et H. Kat, (2002), "The Statistical Properties of Hedge Fund Index Returns and their Implications for Investors", *Journal of Alternative Investments 5*, 25-44.

Brown S. et W. Goetzmann, (1997), "Mutual Fund Styles", *Journal of Financial Economics 43*, 373-399.

Brown S., W. Goetzmann et R. Ibboston, (1999), "Offshore Hedge Funds: Survival & Performance 1989 - 1995", *Yale School of Management Working Paper F-52B*, 38 pages.

Brown S., W. Goetzmann et J. Park, (2000), "Hedge Funds and the Asian Currency Crisis", *Journal of Portfolio Management 26*, 95-101.

Brown S., W. Goetzmann, B. Liang et C. Schwart, (2009), "Estimating Operational Risk for Hedge Funds: The ω-Score", *Financial Analyst Journal 5*, 43-53.

Brys G., M. Hubert et A. Struyf, (2004), "A Robust Measure of Skewness", *Journal of Computational and Graphical Statistics 13*, 996-1017.

Buckley I., D. Saunders et L. Seco, (2008), "Portfolio Optimization when Asset Returns have the Gaussian Mixture Distribution", *European Journal of Operational Research 185*, 1434-1461.

Caballé J. et A. Pomansky, (1996), "Mixed Risk Aversion", *Journal of Economic Theory 71*, 485-513.

Campbell J., A. MacKinlay et A. Lo, (1997), *The Econometrics of Financial Markets*, Princeton University Press, 632 pages.

Carnero M., D. Peña et E. Ruiz, (2007), "Effects of Outliers on the Identification and Estimation of GARCH Models", *Journal of Time Series Analysis 28*, 471-497.

Carrillo S., Hernández N. et L. Seco, (2006-a), "New Families of Distributions fitting L-moments for Modelling Financial Data", *RiskLab Technical Paper*, 42 pages.

Carrillo S., Hernández N. et L. Seco, (2006-b), "A Theoretical Comparison between Moments and L-moments", *RiskLab Technical Paper*, 23 pages.

Cass D. et J. Stiglitz, (1970), "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds", *Journal of Economic Theory 2*, 122-160.

Cenci M. et F. Filippini, (2006), "Portfolio Selection: A Linear Approach with Dual", *Applied Mathematics and Computation 179*, 523-534.

Chamberlain G., (1983), "A Characterization of Distributions that imply Mean-Variance Utility Functions", *Journal of Econmic Theory 29*, 185-201.

Chan L., (1967), "On a Characterization of Distributions by Expected Values of Extreme Order Statistics", *American Mathematical Monthly 74*, 950-951.

Chan L., H. Chen et J. Lakonishok, (2002), "On Mutual Fund Investment Styles", *Review of Financial Studies 15*, 1407-1437.

Chan L., Y. Hamao et J. Lakonishok, (1991), "Fundamentals and Stock Returns in Japan", *Journal of Finance 46*, 1739-1764.

Chang C., T. Pactwa et A. Prakash, (2003), "Selecting a Portfolio with Skewness: Recent Evidence from US, European, and Latin American Equity Markets", *Journal of Banking and Finance 27*, 1375-1390.

Charles A., (2004), "Outliers and Portfolio Optimization", *Banque et Marché 72*, 44-51.

Charles A. et O. Darné, (2005), "Outliers and GARCH Models in Financial Data", *Economics Letters 86*, 347-352.

Charles A. et O. Darné, (2006), "Large Shocks and September 11^{th} Terrorist Attacks: An Intervention Analysis Approach on International Stock Markets", *Economic Modelling 23*, 883-898.

Charnes A., W. Cooper et R. Ferguson, (1955), "Optimal Estimation of Executive Compensation by Linear Programming", *Management Science 1*, 138-151.

Charnes A. et W. Cooper , (1961), *Management Models and industrial Applications of Linear Programming*, Wiley, 467 pages.

Chauveau T. et N. Nalpas, (2008), "Risk Weighted Utility Theory as a Solution to the Equity Premium Puzzle", *Cahiers de la MSE 1999*, 43 pages.

Chen J., H. Hong et J. Stein, (2001), "Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices", *Journal of Financial Economics 61*, 345-381.

Chen S., D. Leung et J. Qin, (2008), "Improving Semiparametric Estimation by using Surrogate Data", *Journal of the Royal Statistical Society 70*, 803-823.

Chen C. et L. Liu, (1993), "Joint Estimation of Model Parameters and Outlier effects in Time Series", *American Statistical Association 88*, 284-297.

Chib S. et E. Greenberg, (1995), "Understanding the Metropolis-Hastings Algorithm", *American Statistician 49*, 327-335.

Chiu H., (2008), "Skewness Preference, Risk Taking and Expected Utility Maximization", Actes du colloque de l'*IESEG*, 27 pages.

Chopra N. et J. Ritter, (1989), "Portfolio Rebalancing and the Turn-of-the-Year Effect", *Journal of Finance 44*, 149-166.

Chopra V. et J. Ziemba, (1993), "The Effects of Errors in Means, Variances and Covariances", *Journal of Portfolio Management 19*, 6-11.

Christodoulakis G. et D. Peel, (2006), "The Relationship between Expected Utility and Higher Moments for Distributions captured by the Gram-Charlier Class", *Finance Research Letters 3*, 273-276.

Christoffersen P., (1998), "Evaluating Interval Forecasts", *International Economic Review 39*, 841-862.

Chu B. et M. Salmon, (2007), "Testing Distributional Assumptions: A L-moment Approach", *Working Paper, Université de Warwick*, 45 pages.

Chunhachinda P., K. Dandapani, K. Hamid et S. Prakash, (1997), "Portfolio Selection and Skewness: Evidence from International Stock Markets", *Journal of Banking and Finance 21*, 143-167.

Cochrane J., (2001), *Asset Pricing*, Princeton University Press, 524 pages.

Cont R., (2001), "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues", *Quantitative Finance 1*, 223-236.

Corielli F. et A. Meucci, (2004), "Pitfalls in linear models for style analysis", *Statistical Methods and Applications 13*, 105-129.

Cornish E. et R. Fisher, (1937), "Moments and Cumulants in the Specification of Distributions", *Review of the International Statistical Institute 5*, 307-320.

Corrado C. et T. Su, (1996), "Skewness and *Kurtosis* in S&P500 Index Returns implied by Option Prices", *Journal of Financial Research 19*, 175-192.

Cottrell M. et J.-C. Fort, (1987), "Étude d'un processus d'auto-organisation", *Annales de l'Institut Henri Poincaré 23*, 1-20.

Crainich D. et L. Eeckhoudt, (2008), "On the Intensity of Downside Risk Aversion", *Journal of Risk and Uncertainty 36*, 267-276.

Cremers K. et A. Petajisto, (2009), "How Active is Your Fund Manager? A New Measure that Predicts Performance", *Review of Financial Studies 22*, 3329-3365.

Cvitanić J., V. Polimenis et F. Zapatero, (2008), "Optimal Portfolio Allocation with Higher Moments", *Annals of Finance 4*, 1-28.

d'Agostino R. et M. Stephens, (1986), *Goodness-of-Fit Techniques*, Marcel Dekker, 576 pages.

da Silva A. et B. de Melo Mendes, (2003), "Value-at-risk and Extreme Returns in Asian Stock Market", *International Journal of Business 8*, 17-40.

Dalén J., (1987), "Algebraic Bounds on Standardized Sample Moments", *Statistics & Probability Letters 5*, 329-331.

Daniel K., M. Grinblatt, S. Titman et R. Wermers, (1997), "Measuring Mutual Fund Performance with Characteristic-based Benchmarks", *Journal of Finance 52*, 1035-1058.

Darolles S., Ch. Gouriéroux et J. Jasiak, (2008), "L-performance with an Application to Hedge Funds", *SSRN Working Paper*, 28 pages.

Das N. et R. Das, (2004), "Hedge Funds Classification Technique using Self-Organized Feature Map Neural Network", *Unpublished Working Paper*, 39 pages.

Davies R., H. Kat et S. Lu, (2006), "Fund of Hedge Funds Portfolio Selection: A Multiple-Objective Approach", *Working Paper, ICMA Center*, 35 pages.

Deboeck G. et T. Kohonen, (1998), *Visual Explorations in Finance with Self-Organizing Maps*, Springer-Verlag, 258 pages.

de Bodt E., M. Cottrell et M. Verleysen, (2002), "Statistical Tools to assess the Reliability of Self-organizing Maps", *Neural Networks 15*, 967-978.

de Bodt E., J. Rynkiewicz et M. Cottrell, (2004), "Some Known Facts about Financial Data", *Neural Networks 17*, 167-183.

de Castro L., (2007), "Fundamentals of Natural Computing: An Overview", *Physics of Life Reviews 4*, 1-36.

de Finetti B., (1940), "Il Problema dei "Pieni"", *Giornale dell' Istituto Italiano degli Attuari 11*, 1-88.

de Roon F., T. Nijman et J. Ter Horst, (2004), "Evaluation Style Analysis", *Journal of Empirical Finance 11*, 29-53.

Dempster A., N. Laird et D. Rubin, (1977), "Maximum Likelihood from Incomplete Data *via* EM Algorithm", *Journal of the Royal Statistical Society 39*, 1-38.

Diacogiannis G., (1994), "Three-parameter Asset Pricing", *Managerial and Decision Economics 15*, 149-158.

diBartolomeo D. et E. Witkowski, (1997), "Mutual Fund Misclassification: Evidence Based on Style Analysis", *Financial Analysts Journal 53*, 32-43.

diCesare G., (2006), "Imputation, Estimation and Missing Data in Finance", *Thèse de doctorat, Université de Waterloo, Canada*, 195 pages.

van Dijk D., P. Franses et A. Lucas, (1999), "Testing for ARCH in the Presence of Additive Outliers", *Journal of Applied Econometrics 14*, 539-562.

Dittmar R., (2002), "Nonlinear Pricing Kernels, *Kurtosis* Preference, and Evidence from the Cross-Section of Equity Returns", *Journal of Finance 57*, 369-403.

Donaldson R. et M. Kamstra, (1997), "An Artificial Neural Network - GARCH Model for International Stock Return Volatility", *Journal of Empirical Finance 57*, 17-46.

Doornik J. et M. Ooms, (2005), "Outlier Detection in GARCH Models", *Tinbergen Institute Working Paper 05-092/4*, 26 pages.

Dorn D., G. Huberman et P. Sengmueller, (2008), "Correlated Trading and Returns", *Journal of Finance 63*, 885-920.

Dounias G. et N. Thomaidis, (2008), "Detecting the Optimal Structure of a Neural Network Model Under Strong Statistical Features in Errors", *SSRN Working Paper*, 18 pages.

Dubauskas G. et D. Teresiené, (2005), "AutoRegressive Conditional Skewness, *Kurtosis* and Jarque-Bera in Lithuanian Stock Market Measurement", *Engineering Economics 5*, 19-24.

Durrleman V., A. Nikeghbali et T. Roncalli, (2000), "Which copula is the right one?", *SSRN Working Paper*, 20 pages.

Eeckhoudt L., Ch. Gollier et Th. Schneider, (1995), "Risk-aversion, Prudence and Temperance", *Economics Letters 48*, 331-336.

Eeckhoudt L. et H. Schlesinger, (2006), "Putting Risk in its Proper Place", *American Economic Review 96*, 280-289.

Efron B. et R. Tibshirani, (1993), *An Introduction to the Bootstrap*, Chapman and Hall, 436 pages.

Ehrgott M., K. Klamroth et C. Schwehm, (2004), "An MCDM Approach to Portfolio Optimization", *European Journal of Operational Research 155*, 752-770.

Elamir E. et A. Seheult, (2003), "Trimmed L-moments", *Computational Statistics & Data Analysis 43*, 299-314.

Elamir E. et A. Seheult, (2004), "Exact Variance Structure of Sample L-moments", *Journal of Statistical Planning and Inference 124*, 337-359.

Eling M., (2006), "Autocorrelation, Bias, and Fat Tails - Are Hedge Funds Really Attractive Investments?", *Journal of Derivatives Use, Trading & Regulation 12*, 28-47.

Embrechts P., (2009), "Linear Correlation and EVT: Properties and Caveats", *Journal of Financial Econometrics 7*, 30-39.

Embrechts P., C. Klüepelberg et Th. Mikosh, (1997), *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, 655 pages.

Embrechts P., A. McNeil et D. Straumann, (1999), "Correlation: Pitfalls and Alternatives", *Risk 12*, 69-71.

Engle R., (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica 50*, 987-1007.

Engle R., (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics 20*, 339-350.

Engle R. et T. Bollerslev, (1986), "Modelling the Persistence of Conditional Variances", *Econometric Reviews 5*, 1-50.

Engle R. et S. Manganelli, (2004), "CAViaR: Conditional Auto-Regressive Value-at-Risk by Regression Quantiles", *Journal of Business and Economic Statistics 22*, 367-381.

Fabozzi F., S. Ortobelli, S. Rachev et H. Shalit, (2006), "Risk Probability Functionals and Probability Metrics Applied to Portfolio Theory", *Working Paper*, 35 pages.

Fama E., (1965), "Random Walks in Stock Market Prices", *Financial Analysts Journal 21*, 55-59.

Fama E., (1996), "Discounting under Uncertainty", *Journal of Business 69*, 415-428.

Fama E. et K. French, (1989), "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics 25*, 23-49.

Fama E. et K. French, (1992), "The Cross-section of Expected Returns", *Journal of Finance 47*, 427-475.

Fama E. et K. French, (1995), "Size and Book-to-Market factors in Earnings and Returns", *Journal of Finance 50*, 131-156.

Fama E. et K. French, (1998), "Value *versus* Growth: The International Evidence", *Journal of Finance 53*, 1975-1999.

Fama E. et K. French, (2007), "The Anatomy of Value and Growth Stock Returns", *Financial Analysts Journal 63*, 96-98.

Fama E. et K. French, (2008), "Dissecting Anomalies", *Journal of Finance 63*, 1653-1678.

Favre L. et J.-A. Galeano, (2002), "Mean-Modified Value-at-Risk Optimization with Hedge Funds", *Journal of Alternative Investment 5*, 21-25.

Fenessey N. et R. Vogel, (1993), "L-moment Diagram should replace Product Moment Diagram", *Water Resources Research 29*, 1745-1752.

Fermanian J.-D. et O. Scaillet, (2002), "Nonparametric Estimation of Copulas for Time Series", *Journal of Risk 5*, 25-54.

Ferson W. et R. Schadt, (1996), "Measuring Fund Strategy and Performance in Changing Economic Condition", *Journal of Finance 51*, 425-461.

Fessant F. et S. Midenet, (2002), "Self-organizing Map for Data Imputation and Correction in Survey", *Neural Computation and Application 10*, 300-310.

Figueroa Garcìa J., D. Kalenatic et C. Lopez Bello, (2008), "Missing Data Imputation in Time Series by Evolutionary Algorithms", paru dans *Advanced Intelligent Computing Theories and Applications - with Aspects of Artificial Intelligence – 4^{th} International Conference on Intelligent Computing*, 275-283.

Filipović D. et N. Vogelpoth, (2007), "A Note on the Swiss Solvency Test Risk Measure", *Review of Financial Studies 20*, 1503-1546.

Fiorentini F. et A. Maravall (1996), "Unobserved Components in ARCH Models: An Application to Seasonal Adjustment", *Journal of Forecasting 15*, 175-201.

Fishburn P., (1979),"On the Foundations of Mean-Variance Analysis", *Theory and Decision 10*, 99-111.

Fleming J., C. Kirby et B. Ostdiek, (2001), "The Economic Value of Volatility Timing", *Journal of Finance 56*, 329-352.

Forgy B., (1965), "Cluster Analysis of Multivariate Data: Efficiency *versus* Interpretability of Classifications", *Biometrics 21*, 768-780.

Fox A., (1972), "Outliers in Time Series", *Journal of the Royal Statistical Society 34*, 350-363.

Franses P. et D. van Dijk, (1999), "Outlier Detection in the GARCH(1,1) Model", *Econometric Institute Research Working Paper EI-9926/RV*, 18 pages.

Franses P. et D. van Dijk, (2000), *Nonlinear Time-series Models in Empirical Finance*, Cambridge University Press, 296 pages.

Franses P. et H. Ghijsels, (1999), "Additive Outliers, GARCH and Forecasting Volatility", *International Journal of Forecasting 15*, 1-9.

French C., (2008), "The Tail that Wags the Hedge Fund Dog", *SSRN Working Paper*, 22 pages.

Frey R. et A. McNeil, (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Time Series: An Extreme Value Approach", *Journal of Empirical Finance 7*, 271-300.

Friedman M. et J. Savage L., (1948), "The Utility Analysis of Choices involving Risk", *Journal of Political Economy 56*, 279-304.

Fromont E., (2006), "L'évalutation du risque et de la performance des Hedge Funds", *Thèse de l'Université de Rennes 1*, 282 pages.

Fry R., V. Martin et C. Tang, (2006), "A New Class of Tests of Contagion: Application to Asian Real Estate and Equity Markets", *Working Paper, Université de Melbourne*, 82 pages.

FTSE, (2002), "The FTSE Global Style Index Series Ground Rules", available on: http://www.ftse.com/Indices/FTSE_Global_Style_Index_Series/Downloads/style_indexrules.pdf, 35 pages.

Fung W. et D. Hsieh, (1997), "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds", *Review of Financial Studies 10*, 275-302.

Fung W. et D. Hsieh, (1999), "Is Mean-variance Analysis Applicable to Hedge Funds?", *Economics Letters 62*, 53-58.

Fung W. et D. Hsieh, (2000), "Performance Characteristics of Hedge Funds and Commodity Funds: Natural *vs.* Spurious Biases", *Journal of Financial and Quantitative Analysis 35*, 291-307.

Fung W. et D. Hsieh, (2001), "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers", *Review of Financial Studies 14*, 313-341.

Fung W. et D. Hsieh, (2002), "Asset-Based Style Factors for Hedge Funds", *Financial Analysts Journal 58*, 16-27.

Gagliardini P. et Ch. Gouriéroux, (2006), "Migration Correlation: Definition and Consistent Estimation", *Journal of Banking and Finance 29*, 865-891.

Galagedera D. et E. Maharaj, (2008), "Wavelet Timescales and Conditional Relationship between Higher-order Systematic Co-moments and Portfolio Returns", *Quantitative Finance 8*, 201-215.

Gallo J. et L. Lockwood, (1997), "Benefits of Proper Style. Classification of Equity Portfolio Managers", *Journal of Portfolio Management 23*, 47-55.

Gamba A. et F. Rossi, (1997), "A Three-moment Based Capital Asset Pricing Model", *Working Paper, Université de Venise*, 16 pages.

Gamba A. et F. Rossi, (1998-a), "A Three-moment Based Portfolio Selection Model", *Rivista di Matematica per le Scienze Economiche e Sociali 20*, 25-48.

Gamba A. et F. Rossi, (1998-b), "Mean-Variance-Skewness Analysis in Portfolio Choice and Capital Markets", *Ricerca Operativa 28*, 5-46.

Garlappi L. et G. Skoulakis, (2008), "Taylor Series Approximations to Expected Utility and Optimal Portfolio Choice", *Working Paper, Université du Texas*, 42 pages.

Gembicki F. et Y. Haimes, (1975), "Approach to Performance and Multiobjective Sensitivity Optimization: The Goal Attainment Method", *Journal of Transactions on Automatic Control 20*, 769-771.

Genest C., B. Rémillard et D. Beaudoin, (2009), "Goodness-of-fit Tests for Copulas: A Review and a Power Study", *Insurance: Mathematics and Economics 44*, 199-213.

Gertler M. et S. Gilchrist, (1994), "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms", *Journal Economics 109*, 309-340.

Gettinby G., C. Sinclair, D. Power et R. Brown, (2006), "An Analysis of the Distribution of Extremes in Indices of Share Returns in the US, UK and Japan from 1963 to 2000", *International Journal of Finance and Economics 11*, 97-113.

Gibson R. et S. Gyger, (2007), "The Style Consistency of Hedge Funds", *European Financial Management, 13*, 287-308.

Gingras D. et K. Adamowski, (1994), "Performance of L-moments and Nonparametric Flood Frequency Analysis", *Canadian Journal of Civil Engineering 21*, 856-862.

Gini C., (1912), "Variabilita e Mutabilita, Contributio allo Studio delle Distribuzionie delle Relazione Statistiche", *tudi Economico-Giuridici della Reale Universita di Cagliari 3*, 3-159.

Glasserman P., (2003), *Monte Carlo Methods in Financial Engineering*, Springer, 610 pages.

Glosten L., R. Jagannathan et D. Runkle, (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance 48*, 1779-1801.

Goetzmann W. et A. Kumar, (2001), "Equity Portfolio Diversification", *NBER Working Paper 8686*, 27 pages.

Goetzmann W., J. Ingersoll et M. Spiegel, (2007), "Portfolio Performance Manipulation and Manipulation-proof Performance Measures", *Review of Financial Studies 20*, 1503-1546.

Golec J. et M. Tamarkin, (1998), "Bettors Love Skewness, Not Risk, at the Horse Track", *Journal of Political Economy 106*, 205-225.

Gonzalo J. et J. Olmo, (2004), "Which Extreme Values are Really Extreme?", *Journal of Financial Econometrics 2*, 349-369.

Gouriéroux Ch. et J. Jasiak, (2001), *Financial Econometrics: Problems, Models, and Methods*, Princeton University Press, 464 pages.

Gouriéroux Ch. et J. Jasiak, (2008), "Dynamic Quantile Models", *Journal of Econometrics 147*, 198-205.

Gouriéroux Ch. et W. Liu, (2006), "Efficient Portfolio Analysis using Distorsion Risk Measures", *CREST Working Paper #2006-17*, 33 pages.

Graham B., (1949), *The intelligent Investor*, Harper Collins, 368 pages.

Grané A. et H. Veiga, (2009), "Wavelet-based Detection of Outliers in Volatility Models", *University of Carlos III, Working Papers ws090403*, 23 pages.

Greenwood J., J. Landweher et N. Natales, (1979), "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form", *Water Ressources Review 15*, 1049-1054.

Gregory-Allen R. et H. Shalit, (1999), "The Estimation of Systematic Risk under Differentiated Risk Aversion: A Mean-Extended Gini Approach", *Review of Quantitative Finance and Accounting 12*, 135-157.

Grossi L., (2004), "Analyzing Financial Time Series through Robust Estimators", *Studies in Nonlinear Dynamics & Econometrics 8*, 18 pages.

Grubbs F., (1969), "Procedures for Detecting Outlying Observations in Samples", *Technometrics 11*, 1-21.

Guastaroba G., R. Mansini et M. Speranza, (2009), "On the Effectiveness of Scenario Generation Techniques in Single-period Portfolio Optimization", *European Journal of Operational Research 192*, 500-511.

Guidolin M. et A. Timmermann, (2005), "Optimal Portfolio Choices under Regime Switching, Skew and *Kurtosis* Preferences", *Working Paper, Federal Reserve Bank of St Louis*, 35 pages.

Guidolin M. et A. Timmermann, (2008), "International Asset Allocation under Regime Switching, Skew and *Kurtosis* Preferences", *Review of Financial Studies 21*, 889-935.

Guinot C., B. Maillet et P. Rousset, (2006), "Understanding and Reducing Variability of SOM Neighbourhood Structure", *Neural Networks 19*, 838-846.

Guttman N., J. Hosking et J. Wallis, (1993), "Regional Precipitation Quantile Values for the Continental US Computed from L-Moments", *Journal of Climate 6*, 2326-2340.

Haas M., (2007), "Do Investors dislike *Kurtosis*?", *Economics Bulletin 7*, 1-9.

Haktanir T., (1997), "Self-determined Probability-weighted Moments Method and its Application to Various Distributions", *Journal of Hydrology 194*, 180-200.

Hamao Y., R. Masulis et V. Ng, (1990), "Correlations in Price Changes and Colatility across International Stock Markets", *Review of Financial Studies 3*, 281-307.

Hamburger H., (1920), "Über eine Erweiterung des Stieljesschen Moment Problems", *Mathematische Zeitschrift 7*, 235-319.

Hampel F., E. Ronchetti, P. Rousseeuw et W. Stahel, (2005), *Robust Statistics: The Approach Based on Influence Functions*, Wiley-Interscience, 502 pages.

Hansen B., (1994), "Autoregressive Conditional Density Estimation", *International Economic Review 35*, 705-730.

Hansen P. et A. Lunde, (2005), "A Forecast Comparison of Volatility Models: does Anything beat a GARCH(1,1)?", *Journal of Applied Econometrics 20*, 873-889.

Hanson R. et C. Lawson, (1974), *Solving Least Square Problems*, Prentice Hall, 340 pages.

Harel O., (2008), "Outfluence - The Impact of Missing Values", *Journal of Model Assisted Statistics and Applications 3*, 161-168.

Harvey C., J. Liechty, M. Liechty et P. Müller, (2003), "Portfolio Selection with Higher Moments", *Working Paper, Université de Duke*, 42 pages.

Harvey C. et N. Shephard, (1996), "The Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns", *Journal of Business and Economic Statistics 14*, 429-434.

Harvey C. et A. Siddique, (1999), "AutoRegressive Conditional Skewness", *Journal of Financial and Quantitative Analysis 34*, 465-487.

Harvey C. et A. Siddique, (2000), "Conditional Skewness in Asset Pricing Tests", *Journal of Finance 55*, 1263-1295.

Hawkins D., (1980), *Identification of Outliers*, Chapman and Hall, 188 pages.

Hawley D., J. Johnson et D. Raina (1990), "Artificial Neural Systems: a new Tool for Financial Decision-making", *Financial Analysts Journal 46*, 63-72.

He C. et T. Teräsvirta, (1999), "Properties of Moments of a Family of GARCH Processes", *Journal of Econometrics 91*, 173-192.

HedgeFund Intelligence, (2008), "HedgeFund Intelligence Report ", disponible sur : http://www.hedgefundintelligence.com/Stub.aspx?StubID=2321, 3 pages.

Heyde C., (1963), "On a Property of the Lognormal Distribution", *Journal of Royal Statistical Society 29*, 392-393.

Hicks J., (1939), *Value and Capital: An Inquiry into some Fundamental Principles of Economic Theory*, Second Edition, Oxford University Press, 340 pages.

Hilderbrand P., (2007), "*Hedge funds* et *prime broker dealer* : éléments de proposition en matière de bonnes pratiques", *Revue de la Stabilité Financière 10*, 73-83.

Hlawitschka W., (1994), "The Empirical Nature of Taylor-series Approximations to Expected Utility", *American Economics Review 84*, 713-719.

Hocquard A., B. Remillard et N. Papageorgiou, (2008), "Replicating the Properties of Hedge Fund Returns", *Journal of Alternative Investments 11*, 49 pages.

Hodge V. et J. Austin, (2004), "A Survey of Outlier Detection Methodologies", *Artificial Intelligence Review 22*, 85-126.

Hoeffding W., (1948), "A Class of Statistics with Asymptotically Normal Distribution", *Annals of Mathematical Statistics 19*, 293-325.

Holt C. et S. Laury, (2002), "Risk Aversion and Incentive Effects", *The American Economic Review 92*, 1644-1655.

Homaifar G. et D. Graddy, (1988), "Equity Yields in Models Considering Higher Moments of the Return Distribution", *Applied Economics 20*, 325-334.

Hong H. et J. Stein, (1999), "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets", *Journal of Finance 54*, 2143-2184.

Hong H. et J. Stein, (2003), "Difference of Opinion, Short-sales Constraints, and Market Crashes", *Review of Financial Studies 16*, 487-525.

Hörmann W., J. Leydold et H. Sak, (2008), "Efficient Risk Simulations for Linear Asset Portfolios", *Working Paper 80, Université de Vienne*, 21 pages.

Hornik K., M. Stinchcombe et H. White, (1989), "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks 2*, 359-366.

ter Horst J., T. Nijman et F. de Roon, (2004), "Evaluating Style Analysis", *Journal of Empirical Finance 11*, 29-53.

Hosking J., (1986), "The Theory of Probability Weighted Moments", *Research Report RC12210, IBM Research Division*, 15 pages.

Hosking J., (1989), "Some Theoretical Results Concerning L-Moments", *Research Report RC14492, IBM Research Division*, 13 pages.

Hosking J., (1990), "L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics", *Journal of the Royal Statistical Society 52*, 105-124.

Hosking J., (2006), "On the Characterization of Distributions by their L-moments", *Journal of Statistical Planning and Inference 136*, 193-198.

Hosking J., (2007), "Some Theory and Practical Uses of Trimmed L-moments", *Journal of Statistical Planning and Inference 137*, 3024-3029.

Hosking J., G. Bonti et D. Siegel, (2000), "Beyond the Log–normal", *Risk 13*, 59-62.

Hosking J. et J. Wallis, (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution", *Technometrics 29*, 339-349.

Hosking J. et J. Wallis, (1997), *Regional Frequency Analysis*, Cambridge University Press, 224 pages.

Hosking J., J. Wallis et J. Wood, (1985), "Estimation of the Generalized Extreme-value Distribution by the Method of Probability-weighted Moments", *Technometrics 27*, 251-261.

Huang C-F. et R. Litzenberger, (1988), *Foundations for Financial Economics*, Prentice Hall, 364 pages.

Huang J., (1989), "Moment Problem of Order Statistics: A Review", *International Statistical Review 57*, 59-66.

Huang J., C. Sialm et H. Zhang, (2008), "Risk Shifting and Mutual Fund Performance", *McCombs Research Paper Series No. FIN-04-08*, 52 pages.

Hughson E., M. Stutzer et C. Yung, (2006), "The Misuse of Expected Returns", *Financial Analysts Journal 62*, 88-96.

Hume D., (1748), *An Enquiry Concerning Human Understanding*, Oxford University Press, 116 pages.

Huisman R., K. Koedijk, C. Kool et F. Palm, (2001), "Tail-index estimates in Small Samples", *Journal of Business and Economic Statistics 19*, 208-216.

Hyvärinen A., J. Karhunen et E. Oja, (2001), *Independent Component Analysis*, John Wiley & Sons, 482 pages.

Hwang S. et S. Satchell, (1999), "Modelling Emerging Market Risk Premium using Higher Moments", *International Journal of Finance and Economics 4*, 271-296.

Ingersoll J., (1975), "Multidimensional Security Pricing", *Journal of Financial and Quantitative Analysis 10*, 785-798.

Ingersoll J., (2008), "Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note", *European Financial Management 14*, 385-390.

Jagannathan R. et T. Ma, (2003), "Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints", *Journal of Finance 58*, 1651-1683.

Jamshidian F. et Y. Zhu, (1996), "Scenario Simulation: Theory and Methodology", *Finance and Stochastics 1*, 43-67.

Jean W., (1971), "The Extension of Portfolio Analysis to Three or More Parameters", *Journal of Financial and Quantitative Analysis 6*, 505-515.

Jean W., (1973), "More on Multidimensional Portfolio Analysis", *Journal of Financial and Quantitative Analysis 8*, 475-490.

Jensen M., (1968), "The Performance of Mutual Funds in the Period 1945-1964", *Journal of Finance 23*, 389-416.

Johansen A. et D. Sornette, (2001), "Large Stock Market Price Drawdowns are Outliers", *Journal of Risk 4*, 55 pages.

Jolliffe I., (1986), *Principal Component Analysis*, Springer-Verlag, 271 pages.

Jondeau E., S. Poon et M. Rockinger, (2007), *Financial Modeling under Non-Gaussian Distributions*, Springer, 542 pages.

Jondeau E. et M. Rockinger, (2003-a), "Conditional Volatility, Skewness and *Kurtosis*: Existence, Persistence and Comovements", *Journal of Economic Dynamic and Control 27*, 1699-1737.

Jondeau E. et M. Rockinger, (2003-b), "How Higher Moments Affect the Allocation of Assets", *Finance Letters 1*, 1-5.

Jondeau E. et M. Rockinger, (2005), "Conditional Asset Allocation under Non-Normality: How Costly is the Mean-Variance Criterion?", *Working Paper, Université de Lausanne*, 42 pages.

Jondeau E. et M. Rockinger, (2006), "Optimal Portfolio Allocation Under Higher Moments", *Journal of the European Financial Management Association 12*, 29-67.

Jondeau E. et M. Rockinger, (2008), "The Economic Value of Distributional Timing", *mimeo*, 47 pages.

Jondeau E. et M. Rockinger, (2009), "The Impact of Shocks on Higher Moments", *Journal of Financial Econometrics 7*, 77-105.

Jorion P., (1986), "Bayes-Stein Estimation for Portfolio Analysis", *Journal of Financial and Quantitative Analysis 21*, 279-292.

Jouanin J.-F., G. Riboullet et T. Roncalli, (2004), "Copula Functions for the Analysis of Dependence Structures", dans *Risk Measures for the 21st Century*, Szego Eds, John Wiley & Sons, 271-319.

Jurczenko E. et B. Maillet, (2001), "The 3-CAPM: Theoretical Foundations and an Asset Pricing Model Comparison in a Unified Framework", dans *Developments in Forecast Combination and Portfolio Choice*, Dunis-Timmermann-Moody Eds, John Wiley & Sons, 239-273.

Jurczenko E. et B. Maillet, (2006-a), "Theoretical Foundations of Asset Allocation and Pricing Models with Higher-order Moments", dans *Multi-moment Asset Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 1-36.

Jurczenko E. et B. Maillet, (2006-b), "The Four-moment Capital Asset Pricing Model: between Asset Pricing and Asset Allocation", dans *Multi-moment Asset Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 113-163.

Jurczenko E., B. Maillet et P. Merlin, (2006), "Hedge Fund Selection with Higher-order Moments: A Non-parametric Mean-Variance-Skewness-*Kurtosis* Efficient Frontier", dans *Multi-moment Asset Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 51-66.

Jurczenko E., B. Maillet et B. Negréa, (2004), "A Note on Skewness and *Kurtosis* Adjusted Option Pricing Models under the Martingale Restriction", *Quantitative Finance 4*, 479-488.

Kahneman D. et A. Tversky, (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica 47*, 263-291.

Kalymon B., (1971), "Estimating Risk in the Portfolio Selection Model", *Journal of Financial and Quantitative Analysis 6*, 559-582.

Kandel S. et R. Stambaugh, (1996), "On the Predictability of Stock Returns: An Asset Allocation Perspective", *Journal of Finance 51*, 385-424.

Karvanen J., (2006), "Estimation of Quantile Mixtures *via* L-moments and Trimmed L-moments", *Computational Statistics and Data Analysis 51*, 947-959.

Kaut M. et S. Wallace, (2006), "Shape-based Scenario Generation using Copulas", *Stochastic Programming E-Print Series 19*, 17 pages.

Keeney R. et H. Raiffa, (1993), *Decision with Multiple Objectives: Preference and Value Tradeoffs*, Cambridge University Press, 589 pages.

Kerstens K., A. Mounir et I. van de Woestyne, (2007), "Geometric Representation of the Mean-Variance-Skewness Portfolio Frontier based upon the Shortage Function", Actes du colloque du X^{th} *EWEPA*, 27 pages.

Kim T. et H. White, (2004), "On More Robust Estimation of Skewness and *Kurtosis*", *Finance Research Letters 1*, 56-73.

Kim T., H. White et D. Stone, (2005), "Asymptotic and Bayesian Confidence Intervals for Sharpe-Style Weights", *Journal of Financial Econometrics 3*, 315-343.

Kimball M., (1990), "Precautionary Saving in the Small and in the Large", *Econometrica 58*, 53-73.

Kimball M., (1992), "Precautionary Motives for Holding Assets", dans *The New Palgrave Dictionary of Money and Finance*, Stockton Press, 158-161.

Kimball M., (1993), "Standard Risk Aversion", *Econometrica 61*, 589-611.

King M. et S. Wadhwani, (1990), "Transmission of Volatility between Stock Markets", *Review of Financial Studies 3*, 5-33.

Kliche D., P. Smith et R. Johnson, (2008), "L-moment Estimators as applied to Gamma Drop Size Distributions", *Journal of Applied Meteorology and Climatology 47*, 3117-3130.

Kliger D. et O. Levy, (2008), "Mood Impacts on Probability Weighting Functions: Large-gamble Evidence", *Journal of Socio-Economics 37*, 1397-1411.

Koenker R. et Z. Xiao, (2006), "Quantile AutoRegression", *Journal of the American Statistical Association 101*, 980-990.

Kofman P. et I. Sharpe, (2003), "Using Multiple Imputation in the Analysis of Incomplete Observations in Finance", *Journal of Financial Econometrics 1*, 216-249.

Kohavi R., (1995), "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", Actes du colloque de *Artificial Intelligence 2*, 1137-1145.

Kohonen T., (1995), *Self-Organizing Maps*, Springer, 521 pages.

Kosowski R., N. Naik et M. Teo, (2007), "Do Hedge Funds deliver Alpha? A Bayesian and Bootstrap Analysis", *Journal of Financial Economics 84*, 229-264.

Kouwenberg R., (2001), "*Scenario* Generation and Stochastic Programming Models for Asset Liability Management", *European Journal of Operational Research 134*, 279-292.

Kraus A. et R. Litzenberger, (1976), "Skewness Preference and the Valuation of Risk Assets", *Journal of Finance 31*, 1085-1099.

Kritzman M., (1994), "What Practitioners need to Know about Time Diversification", *Financial Analysts Journal 50*, 14-18.

Kroll Y., H. Levy et H. Markowitz, (1984), "Mean-Variance *versus* Direct Utility Maximization", *Journal of Finance 39*, 47-61.

Kuenzi D. et X. Shi, (2007), "Asset Based Style Analysis for Equity Strategies: The Role of the Volatility Factor", *Journal of Alternative Investments 10*, 10-24.

Kupiec P., (1995), "Techniques for Verifying the Accuracy of Risk Management Models", *Journal of Derivatives 3*, 73-84.

Lai T., (1991), "Portfolio with Skewness: A Multiple-objective Approach", *Review of Quantitative Finance and Accounting 1*, 293-305.

Lai K., S. Wang et L. Yu, (2006), "Neural Network-based Mean-variance-skewness Model for Portfolio Selection", *Computers and Operations Research 35*, 34-46.

Lajeri-Chaherli F., (2004), "Proper Prudence, Standard Prudence and Precautionary Vulnerability", *Economics Letters 82*, 29-34.

Lakonishok J. et A. Shapiro, (1986), "Systematic Risk, Total Risk and Size as Determinants of Stock Market Returns", *Journal of Banking and Finance 10*, 115-132.

Lakonishok J., A. Shleifer et R. Vishny, (1994), "Contrarian Investment, Extrapolation, and Risk", *Journal of Finance 49*, 1541-1578.

Lattimore P., J. Baker et A. White, (1992), "The Influence of Probability on Risky Choice: a Parametric Examination", *Journal of Economic Behavior and Organization 17*, 377-400.

Ledoit O. et M. Wolf, (2003), "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection", *Journal of Empirical Finance 10*, 603-621.

Ledoit O. et M. Wolf, (2004-a), "A Well-conditioned Estimator for Large-dimensional Covariance Matrices", *Journal of Multivariate Analysis 88*, 365-411.

Ledoit O. et M. Wolf, (2004-b), "Honey, I Shrunk the Sample Covariance Matrix", *Journal of Portfolio Management 30*, 110-119.

Lee J. et M. Verleysen, (2005), "Nonlinear Dimensionality Reduction of Data Manifolds with Essential Loops", *Neurocomputing 67*, 29-53.

Lee S. et S. Maeng, (2003), "Frequency Analysis of Extreme Rainfall using L-moment", *Irrigation and Drainage 52*, 219-230.

Lehman B. et D. Modest, (1987), "Mutual Funds Performance Evaluation: A Comparison of Benchmark and Benchmark of Comparisons", *Journal of Finance 42*, 233-265.

Lendasse A., P. Cardon, V. Wertz, E. de Bodt et M. Verleysen, (2004), "Self-organizing Feature Maps for the Classification of Investment funds", *European Journal of Economic and Social Systems 17*, 183-195.

Lendasse A., E. de Bodt, V. Wertz et M. Verleysen, (2000), "Non-linear Financial Time Series Forecasting - Application to the Bel 20 Stock Market Index", *European Journal of Economic and Social Systems 14*, 81-92.

Lendasse A., E. Oja, O. Simula et M. Verleysen, (2000), "Time Series Prediction Competition: The CATS Benchmark", *Neurocomputing 70*, 2325-2329.

Lendasse A., V. Wertz et M. Verleysen, (2003), "Model Selection with Cross-validations and Bootstraps - Application to Time Series Prediction with Rbfn Models", *Lecture Notes in Computer Science 2714*, 573-580.

León A., G. Rubio et G. Serna, (2005), "AutoRegresive Conditional Volatility, Skewness and *Kurtosis*", *The Quarterly Review of Economics and Finance 45*, 599-618.

Levy H., (1969), "Comment: A Utility Function depending on the First Three Moments", *Journal of Finance 24*, 715-719.

Levy H., (1992), "Stochastic Dominance and Expected Utility: Survey and Analysis", *Management Science 38*, 555-593.

Levy H. et H. Markowitz, (1979), "Approximating Expected Utility by a Function of Mean and Variance", *American Economic Review 69*, 308-317.

Lhabitant F.-S., (2002), *Hedge Funds: Myths and Limits*, John Wiley & Sons, 288 pages.

Lhabitant F.-S., (2004), *Hedge Funds: Quantitative Insights*, John Wiley & Sons, 354 pages.

Liew J. et M. Vassalou, (2000), "Can Book-to-Market, Size and Momentum be Risk Factors that predict Economic Growth?", *Journal of Financial Economics 57*, 221-245.

Lillo F., R. Mantegna et A. Ponzi, (2009), "Market Reaction to Temporary Liquidity Crises and the Permanent Market Impact", à paraître dans *Physics and Society*, 12 pages.

Little R. et D. Rubin, (1987), *Statistical Analysis with Missing Data*, John Wiley & Sons, 408 pages.

Lobosco A. et D. DiBartolomeo, (1997), "Approximating the Confidence Intervals for Sharpe Style Weights", *Financial Analysts Journal 53*, 80-85.

Loistl O., (1976), "The Erroneous Approximation of Excepted Utility by Means of a Taylor's Series Expansion: Analytic and Computational Results", *American Economic Review 66*, 904-910.

Loretan M., (1997), "Generating Market Risk Scenarios using Principal Components Analysis: Methodological and Practical Considerations", *Individual Research Paper, Federal Reserve Board*, 38 pages.

Luenberger D., (1995), *Microeconomic Theory*, McGraw-Hill, 486 pages.

Lux T., (2000), "On Moment Condition Failure in German Stock Returns: An Application of Recent Advances in Extreme Value Statistics", *Empirical Economics 25*, 641-652.

Lux T., (2001), "The Limiting Extremal Behaviour of Speculative Returns: An Analysis of Intra-Daily Data from the Frankfurt Stock Exchange", *Applied Financial Economics 11*, 299-315.

Maillet B. et J.-Ph. Médecin, (2008), "Extreme Volatilities and L-moment Estimations of Tail Indexes", *CES/CNRS Working Paper, Université de Paris-1*, 49 pages.

Maillet B., J.-Ph. Médecin et Th. Michel, (2008), "High Watermarks of Market Risks", à paraître dans *Journal of Mathematical Methods in Economics and Finance*, 17 pages.

Maillet B. et P. Merlin, (2005), "Completing Hedge Fund Missing Net Asset Values using Kohonen Maps and Constrained Randomization", *Artificial Neural Networks: Formal Models and Their Applications, Lecture Notes in Computer Science*, Springer, 923-928.

Maillet B. et P. Merlin, (2007), "Time-series Completion for Robust Asset Allocation and Risk Measurement", *Actes du colloque des Journes de Microconomie Applique 2007*, 29 pages.

Maillet B. et T. Michel, (2003), "An Index of Market Shocks based on Multiscale Analysis", *Quantitative finance 3*, 88-97.

Maillet B. et T. Michel, (2005), "The Impact of the 9/11 Events on the American and French Stock Markets", *Review of International Economics 13*, 597-611.

Maillet B. et P. Rousset, (2003), "Classifying Hedge Funds using Kohonen Map", dans *Connectionist Approaches in Economics and Management Sciences*, Cottrell-Lesage Eds, Kluwer Academic Publisher, 233-259.

Malhotra N., (1987), "Analyzing Market Research Data with Incomplete Information on the Dependent Variable", *Journal of Marketing Research 24*, 74-84.

Malkiel B. et A. Saha, (2005), "Hedge Funds: Risk and Return", *Financial Analysts Journal 61*, 80-88.

Mandelbrot B., (1963), "The Variation of Certain Speculative Prices", *The Journal of Business 36*, 394-419.

Mandelbrot B., (1997), *Fractales, hasard et finance*, Flammarion, 246 pages.

Mandelbrot B., (1999), "A Multifractal Walk down Wall Street", *Scientific American 280*, 70-73.

Maringer D. et P. Parpas, (2009), "Global Optimization of Higher Order Moments in Portfolio Selection", *Journal of Global Optimization 43*, 219-230.

Markowitz H., (1952), "Portfolio Selection", *Journal of Finance 7*, 77-91.

Markowitz H., (1991), "Foundation of Portfolio Theory", *Journal of Finance 46*, 469-477.

Markowitz H., (2006), "De Finetti scoops Markowitz", *Journal of Investment Management 4*, 3-18.

Marschak J., (1938), "Money and the Theory of Assets", *Econometrica 6*, 311-325.

Marschinski R., P. Rossi et M. Tavoni, (2007), "Portfolio Selection with Probabilistic Utility", *Annals of Operational Research 151*, 223-239.

Martellini L. et V. Ziemann, (2007), "Improved Forecasts of Higher-Order Co-moments and Implications for Portfolio Selection", *Working Paper, EDHEC*, 29 pages.

Martins-Filho C. et F. Yao, (2006), "Estimation of Value-at-Risk and Expected Shortfall based on Non-linear Models of Return Dynamics and Extreme Value Theory", *Studies in Nonlinear Dynamics & Econometrics 10*, article 4, 41 pages.

McCulloch B., (2003), "Geometric Return and Portfolio Analysis", *New Zealand Treasury Working Paper 03/28*, 12 pages.

McCulloch W. et W. Pitts, (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics 7*, 115-133.

McLachlan G. et T. Krishnan, (1997), *The EM Algorithm and Extensions*, John Wiley & Sons, 304 pages.

McLeod A. et W. Li, (1983), "Diagnostic Checking ARMA Time-series Models using Squared Residual Autocorrelations", *Journal of Time Series Analysis 4*, 269-273.

Medeiros M., M. McAleer, D. Slottje, V. Ramos et Rey-Maquieira, (2008), "An Alternative Approach to estimating Demand: Neural Network Regression with Conditional Volatility for High Frequency Air Passenger Arrivals", *Journal of Econometrics 147*, 372-383.

Mehra R. et E. Prescott, (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics 15*, 145-161.

Mencía J. et E. Sentana, (2008), "Multivariate Location-scale Mixtures of Normals and Mean-variance-skewness Portfolio Allocation", *CEMFI Working Paper 0805*, 46 pages.

Menezes C. et X. Wang, (2005), "Increasing Outer Risk", *Journal of Mathematical Economics 41*, 875-886.

Meng X. et D. Rubin, (1993), "Maximum Likelihood Estimation *via* the ECM Algorithm", *Biometrika 80*, 267-278.

Merlin P., Sorjamaa A., B. Maillet et A. Lendasse, (2009), "X-SOM and L-SOM: a Nested Approach for Missing Value Imputation", *Actes du colloque de ESANN 2009*, 171-177.

Mervyn M. et S. Wadhwani, (1990), "Transmission of Volatility between Stock Markets", *Review of Financial Studies 3*, 5-33.

Miazhynskaia T., S. Frühwirth-Schnatter et G. Dorffner, (2006), "Bayesian Testing for Nonlinearity in Volatility Modeling", *Computational Statistics and Data Analysis 51*, 2029-2042.

Michaud R., (1989), "The Markowitz Optimization enigma: Is optimized Optimal?", *Financial Analysts Journal 45*, 31-42.

Michaud R., (1998), *Efficient Assset Management: A Practial Guide to Stock Portfolio Optimization*, Oxford University Press, 152 pages.

Michaud O. et R. Michaud, (2008), "Estimation Error and Portfolio Optimization: A Resampling Solution", *Journal of Investment Management 6*, 8-28.

Mikosch T., (2006), "Copulas: Tales and Facts", *Extremes 9*, 3-20.

Mittnik S., M. Paolella et S. Rachev, (2000), "Diagnosing and Treating the Fat Tails in Financial Returns Data", *Journal of Empirical Finance 7*, 389-416.

Mitton T. et K. Vorkink, (2007), "Equilibrium Underdiversification and Preference for the Skewness", *Review of Financial Studies 20*, 1255-1288.

Modarres R., (2008), "Regional Maximum Wind Speed Frequency Analysis for the Arid and Semi-arid Regions of Iran", *Journal of Arid Environments 72*, 1329-1342.

Moisello U., (2007), "On the Use of Partial Probability Weighted Moments in the Analysis of Hydrological Extremes", *Hydrological Process 21*, 1265-1279.

Monarcha G., (2008), "L'analyse dynamique des structures de risque des *Hedge Funds*", *Thèse de l'Université de la Méditéranée*, 287 pages.

Moreno D., P. Marco et I. Olmeda, (2006), "Self-Organizing Maps could improve the Classification of Spanish Mutual Funds", *Journal of Operational Research 174*, 1039-1054.

Morey M. et R. Morey, (1999), "Mutual Fund Performance Appraisals: A Multi-horizon Perspective with Endogenous Benchmarking", *Omega 27*, 241-258.

MSCI, (2007), "MSCI Global Investable Market Value and Growth Indices", disponible sur: http://www.msci.com/methodology/meth_docs/MSCI_May07_GIMIVGMethod.pdf, 30 pages.

Mudholkar G. et A. Hutson, (1998), "LQ-moments: Analogs of L-moments", *Journal of Statistical Planning and Inference 71*, 191-208.

Muler N. et V. Yohai, (2008), "Robust Estimates for GARCH Models", *Journal of Statistical Planning and Inference 138*, 2918-2940.

Nelson D., (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach", *Econometrica 59*, 347-370.

Nelson R., (1998), *An Introduction to Copula*, Springer Verlag, 269 pages.

von Neumann J. et O. Morgenstern, (1944), *Theory of Games and Economic Behavior*, Princeton University Press, 648 pages.

Okunev J., (1988), "A Comparative Study of Gini's Mean Difference and Mean Variance in Portfolio Analysis", *Accounting and Finance 28*, 1-16.

Okunev J., (1989), "Mean Gini Capital Asset Pricing Model: Some Empirical Evidence", *Accounting and Finance 29*, 63-73.

Okunev J., (1990), "An Alternative Measure of Mutual Fund Performance", *Journal of Business Finance and Accounting 17*, 247-264.

Okunev J., (1992-a), "The Generation of Mean Gini Efficient Sets", *Journal of Business Finance and Accounting 18*, 209-218.

Okunev J., (1992-b), "Is the Mean Gini Capital Asset Pricing Model Testable", *Journal of Business Finance and Accounting 19*, 271-277.

Otten R. et D. Bams, (2001), "Statistical Tests For Return-Based Style Analysis", Actes du colloque de l'*EFMA 2001*, 32 pages.

Owen J. et R. Rabinovitch, (1983), "On the Class of Elliptical Distributions and their Applications to the Theory of Portofolio Choice", *Journal of Finance 38*, 745-752.

Pagan A., (1996), "The Econometrics of Financial Markets", *Journal of Empirical Finance 3*, 15-102.

Pandey M., P. van Gelder et J. Vrijling, (2001), "The Estimation of Extreme Quantiles of Wind Velocity using L-moments in the Peaks-over-Threshold Approach", *Structural Safety 23*, 179-192.

Parida B. et D. Moalafhi, (2008), "Regional Rainfall Frequency Analysis for Botswana using L-moments and Radial Basis Function Network", *Physics and Chemistry of the Earth 33*, 614-620.

Pástor Ľ. et R. Stambaugh, (2000), "Comparing Asset Pricing Models: An Investment Perspective", *Journal of Financial Economics 56*, 335-381.

Patton A., (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation", *Journal of Financial Econometrics 2*, 130-168.

Pearson P., (1993), "Application of L-moments to Maximum River Flows", *New Zealand Statististician 28*, 2-10.

Peress J., (2004), "Wealth, Information Acquisition, and Portfolio Choice", *Review of Financial Studies 17*, 879-914.

Perold F., (1984), "Large-scale Portfolio Optimization", *Management Science 30*, 1143-1160.

Pickles A., (2005), "Missing Data, Problems and Solutions", dans *Encyclopedia of Social Measurement*, Kimberly Kempf-Leonard Eds, Elsevier, 689-694.

Pilon P. et K. Adamowski, (1991), "Regional Analysis of Annual Maxima Precipitation using L-moments", *Atmospheric Research 27*, 81-92.

Pilon P. et K. Adamowski, (1992), "The Value of Regional Information to Flood Frequency Analysis using the Method of L-moments", *Canadian Journal of Civil Engineering 19*, 137-147.

Polkovnichenko V., (2005), "Household Portfolio Diversification: A Case for Rank-dependent Preferences", *Review of Financial Studies 18*, 1467-1502.

Popper K., (1934), *The Logic of Scientific Discovery*, Routledge, 544 pages.

Post Th., P. van Vliet et H. Levy, (2008), "Risk Aversion and Skewness Preference", *Journal of Banking and Finance 32*, 1178-1187.

Posthuma N. et P. van der Sluis, (2003), "A Reality Check on Hedge Fund Returns", *VU Research Memorandum 17*, 38 pages.

Pratt J., (1964), "Risk Aversion in the Small and in the Large", *Econometrica 32*, 122-136.

Prelec D., (1998), "The Probability Weighting Function", *Econometrica 66*, 497-527.

Premaratne G. et A. Bera, (2001), "Modeling Asymmetry and Excess *Kurtosis* in Stock Return data", *University of Illinois Working Paper*, 25 pages.

Pressacco F. et P. Serafini, (2007), "The Origins of the Mean-Variance Approach in Finance: Revisiting de Finetti 65 Years Later", *Journal of Decisions in Economics and Finance 30*, 19-49.

Pressacco F. et P. Stucchi, (2000), "Linearity Properties of a Three-moments Portfolio Model", *Decisions in Economics and Finance 23*, 133-150.

Quizon J., H. Binswanger et M. Machina, (1984), "Attitudes toward Risk: Further Remarks", *Economic Journal 94*, 144-148.

Resta M., (2001), "Self-Organizing Maps and Financial Forecasting: An Application", dans *Self-Organizing Neural Networks: Recent Advances and Applications*, Springer-Verlag, Berlin, 185-216.

Riani M., (2004), "Extension of the Forward Search to Time-series", *Studies in Nonlinear Dynamics and Econometrics 8*, 22 pages.

Roh T., (2007), "Forecasting the Volatility of Stock Price Index", *Expert Systems with Applications 33*, 916-922.

Roll R., (1977), "A Critic of the Asset Pricing Theory's Tests: Part I: On Past and Potential Testability of the Theory", *Journal of Financial Economics 4*, 129-176.

Romano J. et A. Siegel, (1986), *Counterexamples in Probability and Statistics*, Wadsworth & Brooks, 303 pages.

Rosenberg B. et M. Houglet, (1974), "Error Rates in CRSP and COMPUSTAT Data Bases and their Implications", *Journal of Finance 29*, 1303-1310.

Rosenblatt F., (1958), "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review 65*, 386-408.

Ross S., (1976), "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory 13*, 341-360.

Rubin D., (1976), "Inference and Missing Data", *Biometrika 63*, 581-592.

Rubin D., (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, 228 pages.

Rubin D., (1996), "Multiple Imputation after 18+ Years", *Journal of the American Statistical Association 91*, 473-489.

Rubinstein M., (1973), "The Fundamental Theorem of Parameter-preference Security Valuation", *Journal of Financial and Quantitative Analysis 8*, 61-69.

Rudin W., (1976), *Principles of Mathematical Analysis*, McGraw-Hill, Third Edition, 342 pages.

Rumelhart D., G. Hinton et R. Williams, (1986), "Learning Representations by Backpropagating Errors", *Nature 323*, 533-536.

Ryoo H., (2007), "A Compact Mean-variance-skewness Model for Large-scale Portfolio Optimization and its Application to the NYSE Market", *Journal of Operational Research Society 58*, 505-515.

Saha A., (1993), "Expo-power Utility: A Flexible Form for Absolute and Relative Risk Aversion", *American Journal of Agricultural Economics 75*, 905-913.

Sahu S., D. Dey et M. Branco, (2003), "A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models", *Canadian Journal of Statistics 31*, 129-150.

Sakata S. et H. White, (1998), "High Breakdown Point Conditional Dispersion Estimation with Application to S&P500 Daily Returns Volatility", *Econometrica 66*, 529-567.

Samad T. et S. Harp, (1992), "Self Organization with Partial Data", *Network 3*, 205-212.

Samuelson P., (1970), "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments", *Review of Economic Studies 37*, 537-542.

Sankarasubramanian A. et K. Srinivasan, (1999), "Investigation and Comparison of Sampling Properties of L moments and Conventional Moments", *Journal of Hydrology 218*, 13-34.

Sarzeaud O. et Y. Stéphan, (2000), "Fast Interpolation using Kohonen Self-Organizing Neural Networks", dans *Theoretical Computer Science: Exploring New Frontiers of Theoretical Informatics*, 126-139.

Schafer J., (1997), *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London, 448 pages.

Schafer J., (1999), "Multiple Imputation: A Primer", *Statistical Methods in Medical Research 8*, 3-15.

Schafer J. et M. Olsen, (1998), "Multiple Imputation for Multivariate Missing-data Problems: A Data Analyst's Perspective", *Multivariate Behavioral Research 33*, 545-571.

Schreiber T., (1998), "Constrained Randomization of Times Series Data", *Physical Review Letter 80*, 2105-2108.

Schweser C., (1978), "Multidimensional Security Pricing: A Correction", *Journal of Financial and Quantitative Analysis 13*, 177-183.

Scott R. et P. Horwath, (1980), "On the Direction of Preference for Moments of Higher Order than the Variance", *Journal of Finance 35*, 915-919.

Semenov A., (2004), "High-Order Consumption Moments and Asset Pricing", *Working Paper, Université de York*, 38 pages.

Serfling R. et P. Xiao, (2007), "A Contribution to Multivariate L-Moments: L-Comoment Matrices", *Journal of Multivariate Analysis 98*, 1765-1781.

Serfling R., (2008), "Survey on (some) Nonparametric and Robust Multivariate Methods", dans *The Finnish Statistical Society 2007*, Read Eds, John Wiley & Sons, 11-41.

Sexton J. et R. Swensen, (2000), "ECM Algorithms that converge at the Rate of the EM", *Biometrika 87*, 651-662.

Sfiridis J., (2005), "Incorporating Higher Moments into Financial Data Analysis", *Working Paper, Université du Connecticut*, 42 pages.

Shalit H. et S. Yitshaki, (1989), "Evaluating the Mean-Gini Approach to Portfolio Selection", *International Journal of Finance 1*, 15-31.

Shalit H. et S. Yitshaki, (2005-a), "The Mean-Gini Efficient Portfolio Frontier", *Journal of Financial Research 28*, 59-75.

Shalit H. et S. Yitshaki, (2005-b), "Capital Market Equilibrium: The Mean-Gini Approach", *Working Paper*, 27 pages.

Shao Q., Y. Chen et L. Zhang, (2008), "An Extension of Three-parameter Burr III Distribution for Low-flow Frequency Analysis", *Computational Statistics & Data Analysis 52*, 1304-1314.

Sharpe W., (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance 19*, 425-442.

Sharpe W., (1988), "Determining the Fund's Effective Asset Mix", *Investment Management Review*, 59-69.

Sharpe W., (1992), "Asset Allocation: Management Style and Performance Measurement", *Journal of Portfolio Management 18*, 7-19.

Siegel A. et A. Woodgate, (2007), "Performance of Portfolios optimized with Estimation Error", *Management Science 53*, 1005-1015.

Sillitto G., (1951), "Interrelations between Certain Linear Systematic Statistics of Sample from any Continuous Population", *Biometrika 38*, 377-382.

Sillitto G., (1964), "Some Relations between Expectations of Order Statistics in Samples of Different Sizes", *Biometrika 51*, 259-262.

Silverman B., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, 176 pages.

Simaan Y., (1993), "Portfolio Selection and Asset Pricing Three Parameter Framework", *Management Science 5*, 568-577.

Simar L. et P. Wilson, (2000), "A General Methodology for Bootstrapping in Non-parametric Frontier Models", *Journal of Applied Statistics 27*, 779-802.

So R. et Y. Tse, (2001), "A Note on International Portfolio Diversification with Short Selling", *Review of Quantitative Finance and Accounting 16*, 311-321.

Sorjamaa A., P. Merlin, B. Maillet et A. Lendasse, (2009), "A Non-linear Approach for Completing Missing Values in Temporal Databases", à paraître dans *European Journal of Economic and Social Systems*, 20 pages.

Stephens M., (1974), "EDF Statistics for Goodness-of-Fit and some Comparisons", *Journal of the American Statistical Association 69*, 730-737.

STOXX, (2007), "Dow Jones STOXX Index Guide", disponible sur: http://www.stoxx.com/download/indices/indexguides/djstoxx_indexguide.pdf, 93 pages.

Sun Q. et X. Yan, (2003), "Skewness Persistence with Optimal Portfolio Selection", *Journal of Banking and Finance 27*, 111-121.

Swinkels L. et P. van der Sluis, (2006), "Return-based Style Analysis with Time-varying Exposures", *European Journal of Finance 12*, 529-552.

Taleb N., (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House, 400 pages.

Taylor J., (2008-a), "Estimating Value-at-Risk and Expected Shortfall using Expectiles", *Journal of Financial Econometrics 6*, 231-252.

Taylor J., (2008-b), "Using Exponentially Weighted Quantile Regression to estimate Value-at-Risk and Expected Shortfall", *Journal of Financial Econometrics 6*, 382-406.

Thompson E., L. Baise et R. Vogel, (2007), "A Global Index Earthquake Approach to Probabilistic Assessment of Extremes", *Journal of Geophysical Research 112*, 12 pages.

Tibiletti L., (2006), "Higher Order Moments and Beyond", dans *Multi-moment Asset Pricing Models*, Jurczenko-Maillet Eds, John Wiley & Sons, 67-78.

Tolikas K., (2008), "Value-at-Risk and Extreme Value Distributions for Financial Returns", *Journal of Risk 10*, 31-77.

Tolikas K. et R. Brown, (2006), "The Distribution of the Extreme Daily Share Returns in the Athens Stock Exchange", *European Journal of Finance 12*, 1-22.

Tolikas K. et G. Gettinby, (2009), "Modelling the Distribution of the Extreme Share Returns in Singapore", *Journal of Empirical Finance 16*, 254-263.

Tolikas K., A. Koulakiotis et R. Brown, (2007), "Extreme Risk and Value-at-Risk in the German Stock Market", *European Journal of Finance 13*, 373-395.

Tolikas K., A. Koulakiotis et G. Dounias, (2006), "An Intelligent Statistical Arbitrage Trading System", dans *Advance in Artificial Intelligence*, Antoniou-Potamias-Spyropoulos-Plexousakis Eds, Springer, 596-599.

Tsiang S.-Ch., (1972), "The Rationale of the Mean-Standard Deviation Analysis, Skewness Preference and the Demand for Money", *American Economic Review 62*, 354-371.

Ulrych T., D. Velis, A. Woodbury et M. Sacchi, (2000), "L-moments and C-moments", *Stochastic Environmental Research and Risk Assessment 14*, 50-68.

Vinod H., (2004), "Ranking Mutual Funds using Unconventional Utility Theory and Stochastic Dominance", *Journal of Empirical Finance 3*, 353-377.

Vogel R. et N. Fennessey, (1993), "L-moment Diagrams Should Replace Product Moment Diagrams", *Water Resources Research 29*, 1745-1752.

Wang Q., (1997), "LH-moments for Statistical Analysis of Extreme Events", *Water Resources Research 33*, 2841-2848.

Wang S., (2003), "Application of Self-organizing Maps for Data Mining with Incomplete Data Sets", *Neural Computation and Application 12*, 42-48.

Wang S. et Y. Xia, (2002), *Portfolio Selection and Asset Pricing*, Springer-Verlag, 200 pages.

Wells J., (2007), "The Rate of Profit as a Random Variable", *Thèse de doctorat, Université Libre de Londres*, 260 pages.

Welsch R. et X. Zhou, (2007), "Application of Robust Statistics to Asset Allocation Models", *Statistical Journal 5*, 97-114.

Whalen T., G. Savage et G. Jeong, (2004), "An Evaluation of the Self-Determined Probability Weighted Moment Method for estimating Extreme Wind Speeds", *Journal of Wind Engineering and Industrial Aerodynamics 92*, 219-239.

White H., T.-H. Kim et S. Manganelli, (2004), "Modeling Autoregressive Conditional Skewness and *Kurtosis* with Multi-Quantile CAViaR", *ECB Working Paper No. 957*, 40 pages.

Wierzbicki A., (1986), "On the Completeness and Constructiveness of Parametric Characterizations to Vector Optimization Problems", *OR Spektrum 8*, 73-87.

Wu G. et R. Gonzalez, (1996), "Curvature of the Probability Weighting Function", *Management Science 42*, 1676-1690.

Yitzhaki S., (2003), "Gini's Mean Difference: A Superior Measure of Variability for Non-Normal Distributions", *International Journal of Statistics 61*, 285-316.

Yu L., S. Wang et K. Lai, (2008), "Neural Network-based Mean-variance-skewness Model for Portfolio Selection", *Computer & Operations Research 35*, 34-46.

Zangari P., (1996), "A VaR Methodology for Portfolios that Include Options", *RiskMetrics Monitor*, 4-12.

Zhang D., (2008), "Non-convex Optimal Portfolio Sets and Constant Relative Risk Aversion", *Journal of Economics and Business 60*, 551-555.

Zhou Y., (2007), "Regime Switching and Stock-Bond Coskewness: Contagion or Flight to Quality", *Working Paper, Université d'Oxford*, 50 pages.

Zumbach G., (2000), "The Pitfalls in Fitting GARCH(1,1) Processes", dans *Advances in Quantitative Asset Management*, Dunis Eds, Kluwer Academic, 179-200.

Annexe :
Articles complémentaires

Annexe 1 :

"Completing Hedge Fund Missing Net Asset Values Using Kohonen Maps and Constrained Randomization"

En collaboration avec Bertrand Maillet.

Paru dans *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005, Lecture Notes in Computer Science*, Springer - Berlin, 923-928.

Completing Hedge Fund Missing Net Asset Values Using Kohonen Maps and Constrained Randomization

Paul Merlin[1] and Bertrand Maillet[2]

[1] A.A.Advisors-QCG (ABN Amro Group), Variances and Paris-1,
(TEAM/CNRS and SAMOS/MATISSE), 72 rue Regnault F-75013 Paris
paul.merlin@malix.univ-paris1.fr
[2] A.A.Advisors-QCG (ABN Amro Group), Variances and Paris-1,
(TEAM/CNRS), 106 bv de l'hôpital F-75647 Paris cedex 13
bmaillet@univ-paris1.fr

Abstract. Analysis of financial databases is sensitive to missing values (no reported information, provider errors, outlier filters…). Risk analysis and portfolio asset allocation require cylindrical and complete samples. Moreover, return distributions are characterised by non-normalities due to heteroskedasticity, leverage effects, volatility feedbacks and asymmetric local correlations. This makes completion algorithms very useful for portfolio management applications, specifically if they can deal properly with the empirical stylised facts of asset returns. Kohonen maps constitute powerful non-linear financial classification tools (see [3], [4] or [6] for instance), following the approach of Cottrell *et al.* (2003), we use a Kohonen algorithm (see [2]), altogether with the Constrained Randomization Method (see [8]) to deal with mutual fund missing Net Asset Values. The accuracy of rebuilt NAV estimated series is then evaluated according to a comparison between the first moments of the series.

1 Introduction

The presence of missing data in the underlying time series is a recurrent problem for asset allocation and risk measure which require to deal with cylindrical and complete samples. Moreover, many financial databases contain missing values. For common stock returns measured at a low frequency, the Gaussian hypothesis is considered as a fairly good approximation, but financial assets such as options can introduce non-linearities and asymmetries to the portfolio returns. Because of the non-normality, symmetric measures of risk as the standard deviation cannot be applied; they do not distinguish between heavy left tails and heavy right tails. Hedge Fund asset return in this sense seems to be very particular. Several empirical studies conclude that many hedge fund index return distributions are not normal and exhibit negative skewness, positive excess kurtosis, and highly significant positive first-order autocorrelation (see [1] for instance). Thus, for hedge fund asset class, higher moments should be taken into account for the analysis. The importance of higher moments of returns, especially the skewness and kurtosis in evaluating portfolio risk and performance has been

W. Duch et al. (Eds.): ICANN 2005, LNCS 3697, pp. 923 – 928, 2005.

already highlighted by a number of authors, proposing and analyzing the inclusion of higher moments in portfolio theory. For illustration in the following, we extracted from the large HFR™ database, a dataset of hedge fund net asset values composed with 49 funds on a 5-year period of 60 monthly values. Note that, at purpose, no missing values are contained in this database.

2 Classical Self-Organized Maps Algorithm

The SOM algorithm is based on the unsupervised learning principle where the training is entirely data-driven and no information about the input data is required (see [5]). The SOM consist of a network, compound in n neurons, units or code vectors organised on a regular low-dimensional grid. If $I = [1, 2, ..., n]$ is the set of the units, the neighbourhood structure is provided by a neighbourhood function Λ defined on I^2 . The network state at time *t* is given by:

$$\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_T(t)] \tag{1}$$

where $\mathbf{m}_i(t)$ is the T-dimensional weight vector of the unit *i*.

For a given state $\mathbf{m}$ and input $\mathbf{x}$, the winning unit $i_w(\mathbf{x}, \mathbf{m})$ is the unit whose weight $\mathbf{m}_{i_w(\mathbf{x}, \mathbf{m})}$ is the closest to the input $\mathbf{x}$.

The SOM algorithm is recursively defined by the following steps:

1. Draw randomly an observation $\mathbf{x}$.
2. Find the winning unit $i_w(\mathbf{x}, \mathbf{m})$ also called the Best Matching Unit (noted BMU) such that:

$$BMU_{t+1} = i_w[\mathbf{x}(t+1), \mathbf{m}(t)] = \underset{\mathbf{m}_i, i \in I}{Argmin} \left\{ \left\| \mathbf{x}(t+1) - \mathbf{m}_i(t) \right\| \right\} \tag{2}$$

where $\|.\|$ is the Euclidian norm.

3. Once the BMU is found, the weight vectors of the SOM are updated so that the BMU and his neighbours are moved closer to the input vector. The SOM update rule is:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon_t \Lambda(BMU, i)[\mathbf{m}_i(t) - \mathbf{x}(t+1)], \forall i \in I \tag{3}$$

where ε_t is the adaptation gain parameter, which is]0,1[-valued, generally decreasing with time. The number of neurons taken into account during the weight updates depends on the neighbourhood function Λ that also generally decreases with time (see [5]).

Figure 1 represents the code vectors obtained using the dataset of hedge funds described above.

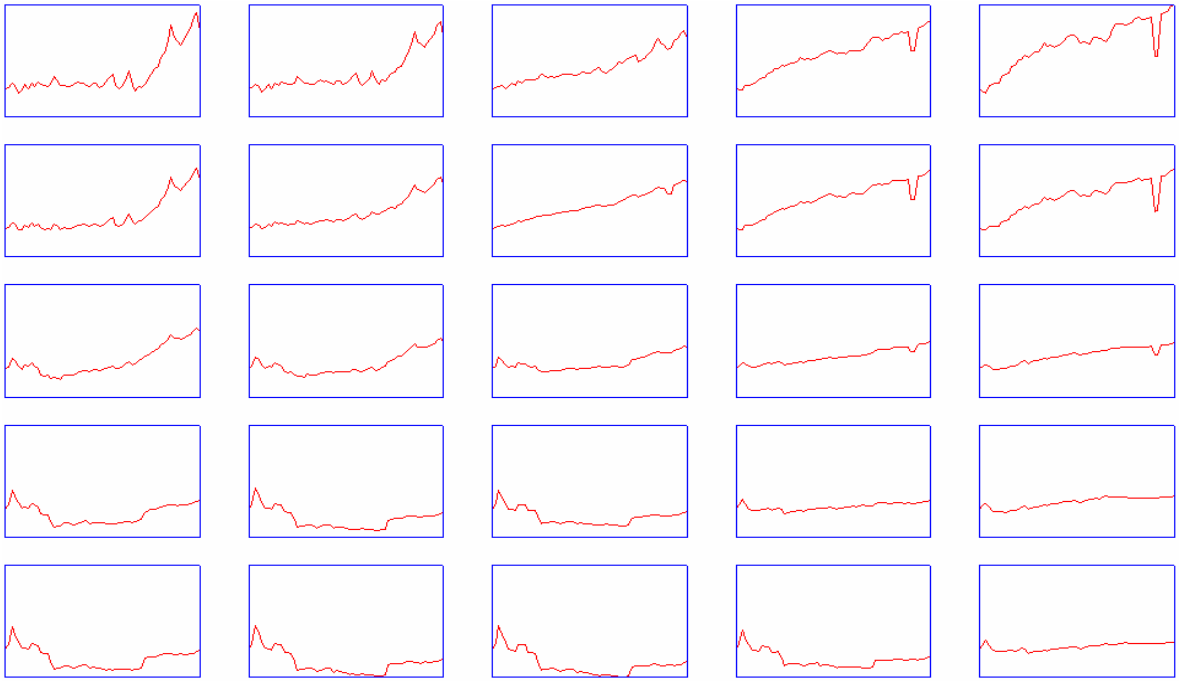


Fig. 1. Representation of Code Vectors on the Kohonen maps

3 Self-Organizing Maps with Partial Data Algorithm

SOM allows for classification of data samples with multiple variables and missing values (see [7]). Cottrell *et al.* (2003) propose an adapted Kohonen algorithm that first clusters the data, and then replaces the missing observations (see [2]). When the SOM algorithm iterates, if a vector $\mathbf{x}$ with missing value(s) is drawn, we consider the subset *NM* of variables which are not missing in vector $\mathbf{x}$. We define a norm on this subset (denotes $\|\cdot\|_M$) that allows us to find the *BMU* (with previous notations):

$$BMU = i_w[\mathbf{x}(t+1), \mathbf{m}(t)] = \underset{i \in I}{Argmin} \left\{ \|\mathbf{x}(t+1), \mathbf{m}(t)\|_M \right\} \tag{4}$$

with:

$$\|\mathbf{x} - \mathbf{m}_i\|_M = \sum_{k \in NM} (\mathbf{x}_k - \mathbf{m}_{i,k})^2$$

where:

$$\begin{cases} \mathbf{x}_k \text{ for } k = [1, ..., T] \text{ denotes the } k^{th} \text{ value of the chosen vector;} \\ \mathbf{m}_{i,k} \text{ for } k = [1, ..., T], \text{ for } i = [1, ..., n] \text{ is the } k^{th} \text{ value of the } i^{th} \text{ code vector;} \\ NM = \text{is the set of the net asset values } \mathbf{x}_k \text{ that are not missing.} \end{cases}$$

Once the Kohonen algorithm has converged, we got some cluster containing our time series. Cottrell *et al.* (2003) first propose to fill the missing values of time-series by the cross-sectional mean of observed values present in the cluster.

4 Combing Self-Organized Maps and Constrained Randomization for Data Completion

Such an approach, when dealing with financial time series, will affect drastically some important statistical properties of the over-all rebuilt dataset. In particular, higher moments (second, third and fourth centred moments), auto-correlations and the correlations with the other time-series are neglected in the analysis. We propose here to combine the Self-Organizing Maps, adapted to the presence of missing values, and the Constrained Randomization algorithm introduced in [8]. This last computational method - initially presented as a specific reshuffling data sampling technique - allows for the simulation of artificial time-series that fulfil given constraints, but are random in other aspects.

The Figure 2 summarizes the proposed procedure for data completion. The first step starts with computing some empirical features of the data (moments of returns in our present case). Then, in parallel, a SOM is run with the non-missing values in the original dataset. Coordinates of Code Vectors in each of *BMU* are then considered as natural first candidates for missing value completion. The constrained randomization, using as constraints some of the empirical features of the data determined at the first step, can then start. If the candidate meets the constraints, then it takes the place of the missing value into the original data; if not, a standard normal residual is drawn, then added to the previous candidates and the test for the constraints starts again. This process lasts until all constraints are fulfilled and all missing values replaced.

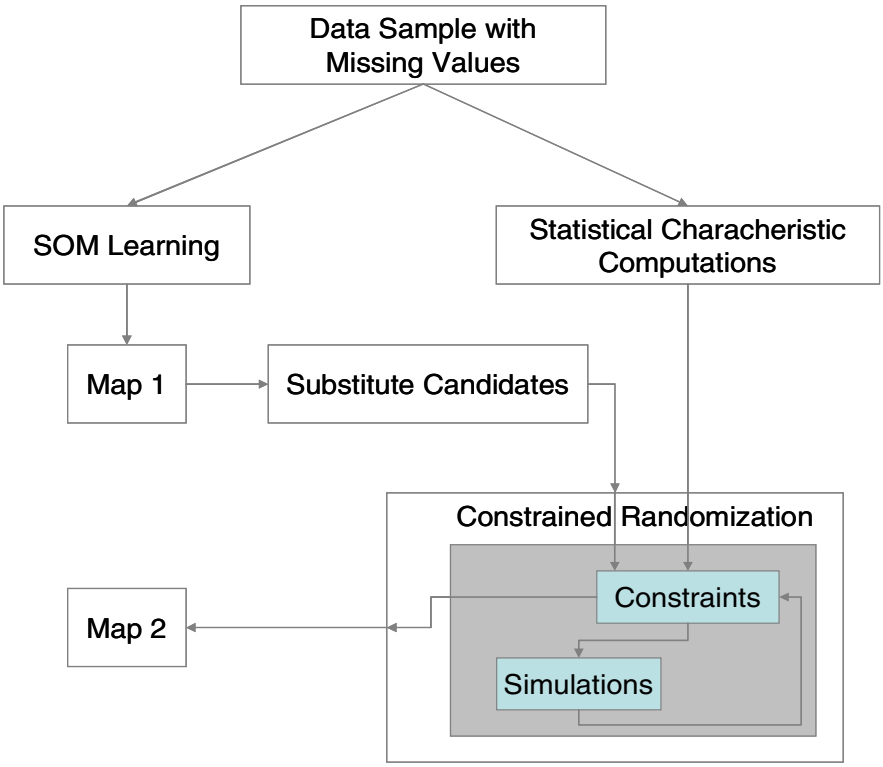


Fig. 2. Representation of the Scheme when Mixing Self-Organizing Maps and Constrained Randomization in Data Completion

5 Empirical Illustrations

Table 1 and Table 2 hereafter summarize the mean properties of the errors in moments when - respectively - using the adapted Kohonen algorithm alone and the two-step procedure presented in this article. As a first remark, we can note that - with no surprise - the addition of a Constrained Randomization procedure allows to recover missing values that more in line with the statistical characterization of the original series, as indicated by the comparison of Table 1 and Table 2. As the second remark, this is true in our example for all (reasonable) level of missing values in the original database. Finally, in this example, the improvement of accuracy regarding the moments is between 11% and 46%, the mean improvement is of order of 27%.

Table 1. Mean Errors on Moments when using the adapted SOM algorithm for Missing Values- fifty draws

	Absolute Error (in %) after Completion *via* Kohonen Maps			
Missing Values (in %)	Mean	Variance	Skewness	Kurtosis
5.00	2.99	6.96	11.55	9.81
10.00	5.38	12.71	20.71	17.71
15.00	7.47	17.82	27.20	24.21
20.00	9.15	23.51	31.90	29.74
25.00	10.85	28.04	36.89	34.40
30.00	13.91	34.04	40.44	40.85
35.00	15.59	38.83	45.62	45.83
40.00	19.57	42.81	45.04	47.97
45.00	21.62	47.01	54.13	52.88
50.00	23.60	54.77	61.67	62.21

Source: HFR™; Monthly Net Asset Values (12/1999-12/2004). Computations from the authors.

Table 2. Mean Errors on Moments when using the adapted SOM algorithm for Missing Values and Constrained Randomization - fifty draws

	Absolute Error (in %) after Completion *via* Kohonen Maps Combined with Constrained Randomization			
Missing Values (in %)	Mean	Variance	Skewness	Kurtosis
5.00	2.65	4.63	6.26	5.59
10.00	4.56	8.95	12.99	10.52
15.00	6.13	13.39	17.72	15.43
20.00	7.26	17.69	21.55	20.13
25.00	8.55	22.08	27.26	24.82
30.00	9.42	26.65	30.79	30.38
35.00	10.43	30.04	34.07	34.62
40.00	12.02	33.48	36.22	36.82
45.00	12.98	37.42	41.45	41.47
50.00	14.43	43.82	48.51	49.56

Source: HFR™; Monthly Net Asset Values (12/1999-12/2004). Computations from the authors.

For a more illustrative example, let us suppose a 17% annualized return fund. We destruct artificially 5%, 20% and 50% of the time series, at a 5% level of missing values, both methodologies get the same result with 0.5 points error on annualized return estimated (the annualized return estimated is between 16.5% and 17.5%). At a 20% level of missing value, the difference between the two methodologies is more observable: 1.5 points for a completion with Kohonen maps *versus* 1 point for a completion with Kohonen Maps combined with Constrained Randomization Method. At a 50% level of missing value, the difference becomes explicit: 4 points for a completion with Kohonen Maps *versus* 2 points for a completion with Kohonen Maps combined with Constrained Randomization Method.

6 Conclusion

The presented method for data completion uses SOM description of the data as the starting point for a constrained randomization. The main interest of the technique can be found in the fact that some of the important empirical features of the input are respected during the rebuilding process of missing observations. Specifically higher moments, whose accuracy of estimations are crucial in some financial applications, are taken into account when substitutions. Moreover, one can easily think about some generalizations of the proposed algorithm, adding for instance some features under studies into the constraints of the so-called Constrained Randomization procedure, such as local correlation structure or tail of the density focuses, depending on what is the final aim of the financial applications (asset allocation or risk management). One may also think about the robustness of the algorithm, namely specifying robust estimators in the constraints and allowing for data resampling when building the Kohonen Maps (see [4]).

References

1. Agarwal, V., Naïk, N.: Multi-period Performance Persistence Analysis of Hedge Funds, Journal of Financial and Quantitative Analysis 35 (2000), 327-342.
2. Cottrell, M., Ibbou, S., Letrémy, P.: *Traitement des données manquantes au moyen de l'algorithme de Kohonen*, in french in Proceedings of the tenth ACSEG Conference (2003), 12 pages.
3. Cottrell, M., de Bodt, E., Grégoire, P.: Financial Application of the Self-Organizing Map, Proceedings of EUFIT'98 (1), Verlag Mainz, (1998), 205-209.
4. De Bodt, E., Cottrell, M.: Bootstrapping Self-Organizing Maps to Assess the Statistical Significance of Local Proximity, European Symposium on Artificial Neural Networks (2000), 245-254.
5. Kohonen, T.: Self-Organizing Maps, Springer, Berlin (1995), 362 pages.
6. Maillet, B., Rousset, P.: Classifying Hedge Funds using Kohonen Map, in Connectionist Approaches in Economics and Management Sciences, Series in Advances in Computational Management Science, Vol. 6, Cottrell-Lesage (Eds), Kluwer Academic Publisher, 2003, 233-259
7. Samad, T., Harp, S.: Self Organization with Partial Data, Network 3, (1992), 205-212.
8. Schreiber, T.: Constrained Randomization of Times Series Data, Physical Review Letter 80 (10) (1998), 2105-2108.

Annexe 2 :

"Robust SOM for Realistic Data Completion"

En collaboration avec Bertrand Maillet et Patrick Rousset.

Paru dans les actes du colloque du 5^{th} *Workshop on Self-Organizing Maps, Paris*, 2005, 371-378.

ROBUST SOM FOR REALISTIC DATA COMPLETION

Bertrand Maillet[1], Paul Merlin[2], Patrick Rousset[3]
[1]A.A.Advisors-QCG (ABN Amro Group), Variances and Paris-1 (TEAM/CNRS),
106 bv de l'hôpital F-75647 Paris cedex 13. France
bertrand.maillet@univ-paris1.fr
[2]A.A.Advisors-QCG (ABN Amro Group), Variances and Paris-1 (TEAM/CNRS and SAMOS/MATISSE), 72 rue Regnault F-75013 Paris. France
paul.merlin@malix.univ-paris1.fr
[3] CEREQ, 10 place de la Joliette
F-13567 Marseille. France
rousset@cereq.fr

Abstract – *Self-Organizing Maps aims ideally to group homogeneous individuals, highlighting a neighbourhood structure between classes in a chosen network. Recent approaches propose to exploit the homogeneity of the underlying classes for data completion purposes (see [2]). The aim of this paper is two-fold. First, we present and slightly modified two complementary approaches in completing the stochastic method proposed by Rousset and Maillet [11] based on bootstrap process for increasing the reliability of the induced neighbourhood structure and, second, we use the induced Robust Map of the last approach for data completion, generalising the results by Merlin and Maillet [9] with robust statistics of the moments of the series. An empirical illustration of this new completion scheme is finally provided based on a sample of Hedge Fund Net Asset Values.*

Key words – Self-Organizing Maps, Missing Value, Bootstrap, Constrained Randomization, Neighbourhood Structure

1 Introduction

The presence of missing data in the underlying time series is a recurrent problem when dealing with databases. Moreover, many financial databases contain missing values. For common stock returns measured at a low frequency, the Gaussian hypothesis is considered as a fairly good approximation, but financial assets such as options can introduce non-linearities and asymmetries to the portfolio returns. Because of the non-normality, symmetric measures of risk as the standard deviation cannot be applied; they do not distinguish between heavy left tails and heavy right tails. Hedge Fund asset return in this sense seems to be very particular. Several empirical studies conclude that many hedge fund index return distributions are not normal and exhibit negative skewness, positive excess *kurtosis*, and highly significant positive first-order autocorrelation (see [1] for instance). Thus, for the hedge fund asset class, higher moments should be taken into account for the analysis. The importance of higher moments of returns, especially the skewness and *kurtosis* in evaluating portfolio risk and performance has been already highlighted by a

number of authors (see [7]), proposing and analyzing the inclusion of higher moments in portfolio theory. For illustration in the following, we extracted from the large HFR™ database, a dataset of hedge fund net asset values composed with 149 funds on a 10-year period of 120 monthly values. Note that, at purpose, no missing values are contained in this database.

2 Classical Self-Organized Maps Algorithm

The SOM algorithm is based on the unsupervised learning principle where the training is entirely data-driven and no information about the input data is required (see [8]). The SOM consist of a network, compound in n neurons, units or code vectors organised on a regular low-dimensional grid. If $I = [1, 2, ..., n]$ is the set of the units, the neighbourhood structure is provided by a neighbourhood function Λ defined on I^2. The network state at time t is given by:

$$\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_T(t)] \quad (1)$$

where $\mathbf{m}_i(t)$ is the T-dimensional weight vector of the unit i.

For a given state **m** and input **x**, the winning unit $i_w(\mathbf{x},\mathbf{m})$ is the unit whose weight $\mathbf{m}_{i_w(\mathbf{x},\mathbf{m})}$ is the closest to the input **x**.

The SOM algorithm is recursively defined by the following steps:

1. Draw randomly an observation **x**.
2. Find the winning unit $i_w(\mathbf{x},\mathbf{m})$ also called the Best Matching Unit (noted *BMU*) such that :

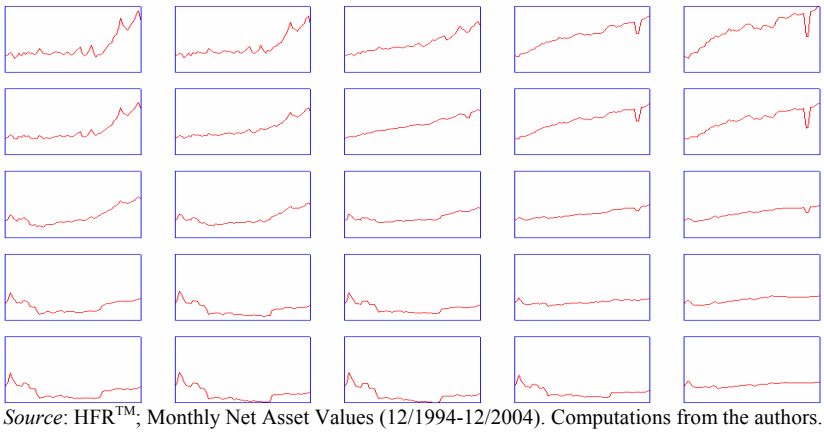
$$BMU_{t+1} = i_w[\mathbf{x}(t+1), \mathbf{m}(t)] = \underset{\mathbf{m}_i, i \in I}{Argmin} \{ \| \mathbf{x}(t+1) - \mathbf{m}_i(t) \| \} \quad (2)$$

where $\|.\|$ is the Euclidian norm.

3. Once the BMU is found, the weight vectors of the SOM are updated so that the BMU and his neighbours are moved closer to the input vector. The SOM update rule is :

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon_t \Lambda(BMU, i)[\mathbf{m}_i(t) - \mathbf{x}(t+1)], \forall i \in I \quad (3)$$

where ε_t is the adaptation gain parameter, which is]0,1[-valued, generally decreasing with time. The number of neurons taken into account during the weight updates depends on the neighbourhood function Λ that also generally decreases with time (see [3]).



Source: HFR™; Monthly Net Asset Values (12/1994-12/2004). Computations from the authors.

Figure 1: Representation of Code Vectors on the Kohonen Maps

3 Building a Robust Map

When SOM are used in classification, the algorithm is applied to the complete database that is generally a sample of some unknown stationary distribution. A first concern refers to the question of the stability of the SOM solution (specifically the neighbourhood organisation) to changes in the sample and to contamination by large outliers. A second concern regards the stability to the data presentation order and the initialisation. For limiting the dependence of the outputs to the original data sample and to the arbitrary choices within an algorithm, it is common to use a bootstrap process with a re-sampling technique (see [4]). Here, this idea is applied to the SOM algorithm, when estimating an empirical probability for any pair of individuals to be neighbours in a map. This probability is estimated by the number of times the individuals have been neighbours at ray 1 when running several times the same SOM algorithm using re-sampled data series (see Figure 2). In the following, we call *P* the matrix containing empirical probabilities for two individuals to be considered as neighbours at the end of the classification. Following Rousset and Maillet [11], the algorithm uses only individuals in the given re-sampled set of individuals (representing 60% or so of the original population). We generalize the previous approach by adding a drawing without replacement in the original series of most the observations (around 60%) for each individuals. At the end of the first step, the left incomplete individuals are classified using computed distances to the code vectors. Thus, at each step, the table of empirical probabilities concerns all individuals in the original dataset, even if only a partial part of them have been used within the algorithm.

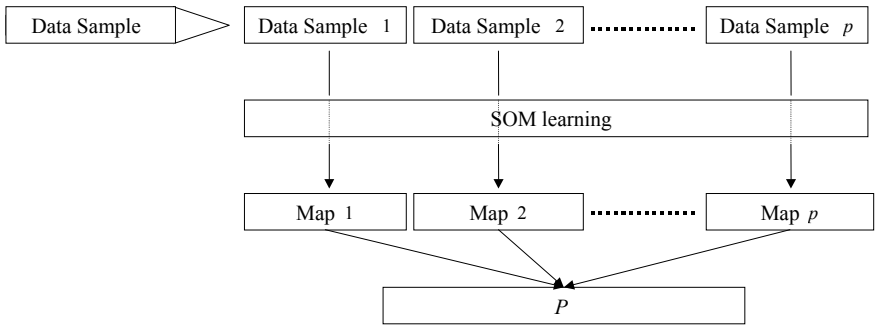


Figure 2: Step1, bootstrap process for building the table *P* of the individual's empirical probabilities to be neighboured one-to-one.

When the matrix *P* is built, the first step is over. In the second step (see Figure 3), the SOM algorithm is also executed several times, but without re-sampling. For any map M_i, we can build the table P_{M_i}, similar to previous one, in which values are 1 for a pair of neighbours and 0 for others. Then, using the Frobenius norm, we can compute the distance between both neighbourhood structures, defined respectively at the end of step 1 (re-sampling the data) and step 2 (computing several maps with the original data). The Robust Map selected, called hereafter R-Map for the sake of simplicity, is the one which minimizes the distance between the two neighbourhood structures as follows:

$$\text{R - Map} = \underset{M_i, i \in I}{Argmin} \left\{ \left\| P - P_{M_i} \right\|_{Frob} \right\} \quad \textbf{(4)}$$

where $\|.\|_{Frob}$ is the Frobenius norm, that is:

$$\|\mathbf{A}\|_{Frob} = {}^{1}\!/_{n^2} \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{[i,j]}^2} \tag{5}$$

with n the dimension of the square matrix $\mathbf{A}$, whose elements are $a_{[i,j]}, \forall (i,j) \in I^2$.

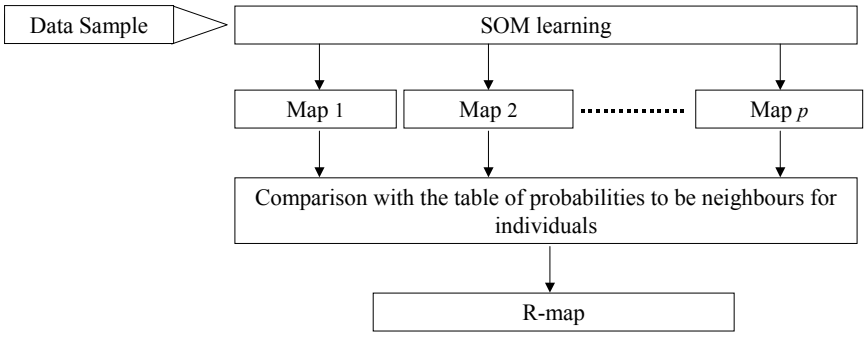


Figure 3: Step2, get the R-Map by selecting the map whose neighbourhood structure is the closest to the empirical probability table *P* obtained at step 1.

4 Robust Self-Organizing Maps with Partial Data Algorithm

SOM allow for classification of data samples with multiple variables and missing values (see [12]). Cottrell *et al.* (2003) propose an adapted Kohonen algorithm that first clusters the data, and then replaces the missing observations (see [2]). When the SOM algorithm iterates, if a vector **x** with missing value(s) is drawn, we consider the subset *NM* of variables which are not missing in vector **x**. We define a norm on this subset (denotes $\|.\|_M$) that allows us to find the *BMU* (with previous notations):

$$BMU = i_w[\mathbf{x}(t+1), \mathbf{m}(t)] = \underset{i \in I}{Argmin}\left\{ \|\mathbf{x}(t+1) - \mathbf{m}_i(t)\|_M \right\} \tag{6}$$

with:

$$\|\mathbf{x} - \mathbf{m}_i\|_M = \sum_{k \in NM} (\mathbf{x}_k - \mathbf{m}_{i,k})^2$$

where:

$$\begin{cases} \mathbf{x}_k \text{ for } k = [1, ..., T] \text{ denotes the } k^{th} \text{ value of the chosen input vector;} \\ \mathbf{m}_{i,k} \text{ for } k = [1, ..., T], \text{ for } i = [1, ..., n] \text{ is the } k^{th} \text{ value of the } i^{th} \text{ code vector;} \\ NM \text{ is the set of the net asset values } \mathbf{x}_k \text{ that are not missing.} \end{cases}$$

Once the Kohonen algorithm has converged, we got some cluster containing our time series. Cottrell *et al.* (2003) first propose to fill the missing values of time-series by the cross-sectional mean of observed values present in the cluster. It is then straightforward to adapt the previous algorithm with the use of the Robust Map defined in the previous sub-section.

5 Combining the Robust Self-Organizing Maps and a Constrained Randomization Procedure for Data Completion

The previous approach will nevertheless affect drastically some important statistical properties of the over-all rebuilt dataset. In particular, higher moments (second, third and fourth centred moments) are neglected in the analysis. Merlin and Maillet [9] propose to combine the Self-Organizing Maps, adapted to the presence of missing values, and the Constrained Randomization algorithm introduced in [13]. This last computational method - initially presented as a specific reshuffling data sampling technique - allows for the simulation of artificial time-series that fulfil given constraints, but are random in other aspects.

The Figure 4 summarizes the proposed procedure for data completion. The first step starts with computing some empirical features of the data (moments of returns in our present case). Then, in parallel, the Robust Map is determined only using the non-missing values in the original dataset. Coordinates of Code Vectors in each unit of the Robust Map are then considered as natural first candidates for missing value completion (see [2]). The constrained randomization, using as constraints some of the empirical features of the data determined at the first step, can then start. If the candidate meets the constraints, then it takes the place of the missing value into the original data; if not, a residual noise is drawn[1], and added to the previous candidates then the test for the constraints starts again.

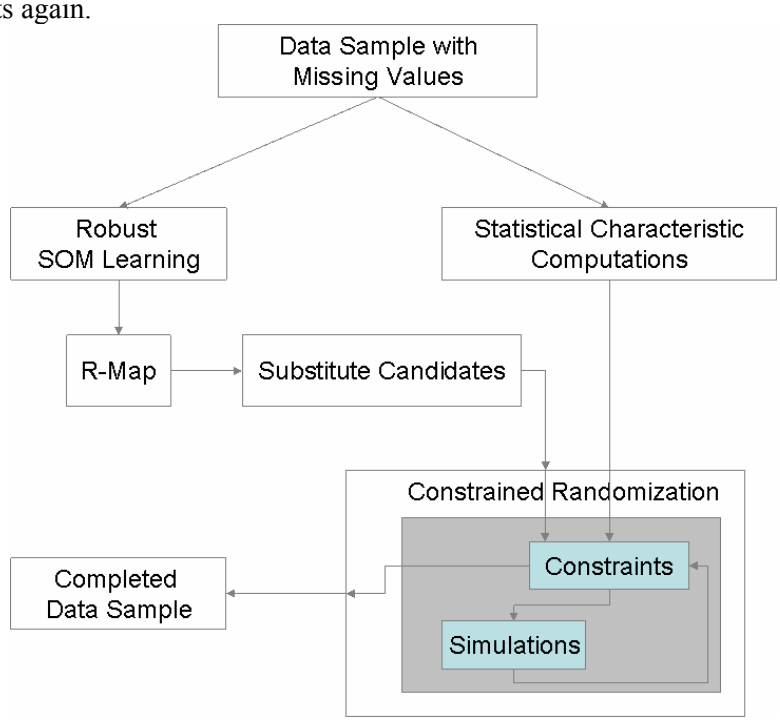


Figure 4: Representation of the Scheme when Mixing Robust Self-Organizing Maps and Constrained Randomization in Data Completion.

In comparison with Merlin and Maillet [9], who use the (simple) sample empirical counterparts of the four first moments of the originals series, we use here *L*-moments as defined in [6] in the procedure of Constrained Randomization such as:

[1] Since our application focuses on financial variables, the noise is drawn from a central Skew Student's t-distribution introduced in [5], with five degrees of freedom as mentioned in [10].

$$\left\| \mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_{NM}) \right\|_{Frob} < \varepsilon \tag{7}$$

where $\|.\|_{Frob}$ is the Frobenius norm, $\mathbf{L}(.)$ is the first four *L*-moments matrix, $\mathbf{x}_{NM}$ is the original series (without missing value), and $\mathbf{x}$ the ultimate rebuilt complete dataset.
Indeed, *L*-moments are some linear combinations of order statistics b_i, $i = [1, ..., r]$ that have simple interpretations as measures of the location, dispersion and shape of the data sample. The have also the advantage of being more stable and less sensitive to outliers. More precisely, the first *L*-moments are defined by:

$$\begin{cases} l_1 = b_0 \\ l_2 = 2\,b_1 - b_0 \\ l_3 = 6\,b_2 - 6\,b_1 + b_0 \\ l_4 = 20\,b_3 - 30\,b_2 + 12\,b_1 - b_0 \end{cases} \tag{8}$$

where:

$$\begin{cases} b_0 = T^{-1} \sum_{j=1}^{T} X_j \\ b_r = T^{-1} \sum_{j=r+1}^{T} \dfrac{(j-1)(j-2)...(j-r)}{(n-1)(n-2)...(n-r)} X_j \end{cases}$$

and with $[X_1, X_2, ..., X_T]$ is the set of observations sorted by increasing order.

The algorithm of completion thus starts by filling missing observations with the corresponding value of the Code Vector associated to the individuals on the non-missing value periods. If the new rebuilt value meets the conditions of equation (7) then the algorithm stops; otherwise a random *alea* is drawn from a Skew Student's t-distribution with five degrees of freedom, and is added to the previous substitute. If the new rebuilt value meets the conditions of equation (7), then the algorithm stops and the database is completed; if not, another draw is made and added to the corresponding value of the Code Vector; and so on until the condition in equation (7) is fulfilled.

6 An Empirical Illustration

Table 1 hereafter summarizes the mean properties of the errors in *L*-moments when using respectively the two-step procedure and the algorithm presented by Cottrell *et al.* (2003) in [2] (in brackets) in the worst case[2]. As a general remark, we can note that - with no surprise - the addition of the Constrained Randomization procedure to the a R-Map determination procedure allows to recover missing values that are more in line with the statistical characterization of the original series. The error terms are very low in general (under 1% for the first and second *L*-moments), even for unrealistic high rates of missing data. Note also that errors in the higher *L*-moments are always lower than the rate of deletion. Finally, in this example, the improvement of

[2] The worst case corresponds to the Map obtained during the second step of the R-Map construction which maximizes the distance between its neighbourhood structure and the *P* matrix obtained during the first step of the R-Map construction. It allows to compare the two methodologies in the sense that the algorithm provide in [2] can come up with some large errors in case of bad luck.

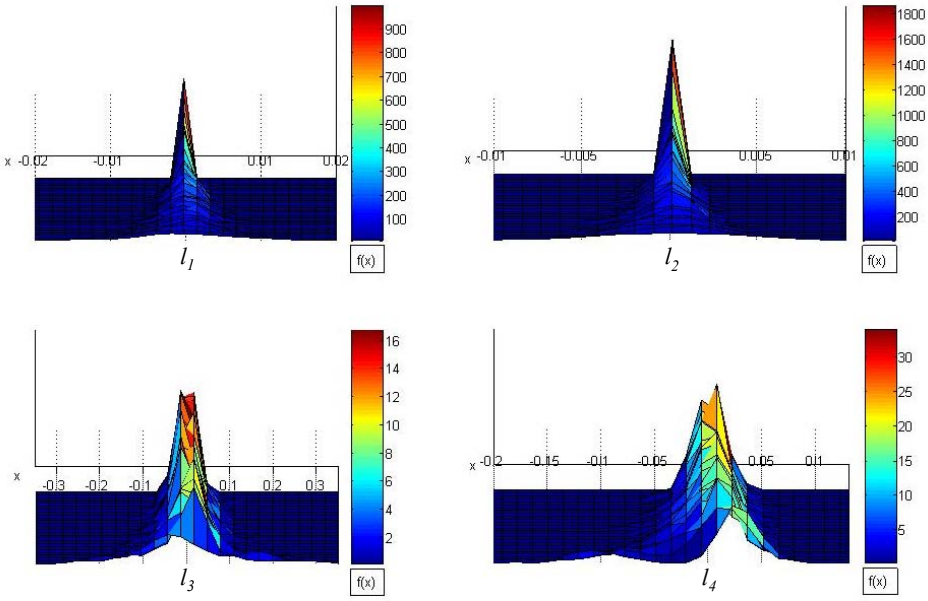
the accuracy regarding the *L*-moments is between 80% and 90% when comparing the two-step procedure and the original procedure worst case.

	Absolute Error (in %) after Completion *via* Robust Kohonen Maps combined with Constrained Randomization							
Missing Values	Mean		Variances		Skewness		*Kurtosis*	
5	**0.01**	[0.07]	**0.01**	[0.06]	**0.25**	[1.56]	**0.16**	[1.23]
10	**0.02**	[0.11]	**0.01**	[0.10]	**0.33**	[2.45]	**0.19**	[2.05]
15	**0.02**	[0.14]	**0.01**	[0.15]	**0.45**	[3.09]	**0.27**	[2.58]
20	**0.03**	[0.16]	**0.02**	[0.20]	**0.46**	[3.44]	**0.32**	[3.27]
25	**0.03**	[0.19]	**0.02**	[0.24]	**0.54**	[4.33]	**0.37**	[3.63]

Source: HFR™; Monthly Net Asset Values (12/1994-12/2004). Computations from the authors.

Table 1. Mean Errors on *L*-moments when using respectively the adapted Robust SOM algorithm for Missing Values and Constrained Randomization (in bold) and the SOM algorithm for Missing Values presented in [2] (in brackets) – for fifty draws.

To illustrate the accuracy of the estimation procedure, we present hereafter the non-parametric empirical densities of the first four *L*-moments for each fund obtained for fifty trials of the complete algorithm for a 20 % deletion level.



Source: HFR™; Monthly Net Asset Values (12/1994-12/2004). Computations from the authors.

Figure 5: Representation of the densities of the first four *L*-moments of the 149 fund returns obtained for a 20% deletion level after fifty draws (centred on the *L*-moment estimates before completion). Centred *L*-moments are on the x-axis, the different funds are on the y-axis, whilst the empirical estimations of the densities appear on the z-axis.

7 Conclusion

The presented method for data completion uses SOM description of the data, in a modified robust version presented in [11] as the starting point for a constrained randomization presented in [13] revised in this paper for being less sensitive to outliers and noise in the data. The main interest of the technique can be found in the fact that some of the important empirical features of the input are respected during the rebuilding process of missing observations. Specifically higher moments,

whose accuracy of estimations are crucial in some financial applications, are taken into account when substitutions. Moreover, one can easily think about some generalizations of the proposed algorithm, adding for instance some features under studies into the constraints of the so-called Constrained Randomization procedure, such as local correlation structure or tails of the density focuses, depending on what is the final purpose and uses of the completed database. Empirical applications such asset allocation or risk management could take benefit of such technique in the sense that their efficiency crucially depends on the reliability of the financial data characteristics.

References

[1] Agarwal, V., Naïk, N. (2000), Multi-period Performance Persistence Analysis of Hedge Funds, *Journal of Financial and Quantitative Analysis*, **vol. 35**, p. 327-342.

[2] Cottrell, M., Ibbou, S., Letrémy, P. (2003), Traitement des données manquantes au moyen de l'algorithme de Kohonen, in French, *in Proceedings of the tenth ACSEG Conference*, 12 pages.

[3] Cottrell, M., Fort, J.C., Pages, G. (1998), Theoretical Aspects of the SOM Algorithm, *Neurocomputing*, **vol. 21,** p. 119-138.

[4] Efron, B.,Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

[5] Hansen, B. (1994), Autoregressive Conditional Density Estimation, *International Economic Review*, **vol. 35**, p. 705-730.

[6] Hosking, J. (1990), *L*-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, *Journal of the Royal Statistical Society*, **vol. 52(2)**, p.105-124.

[7] Jondeau, E., Rockinger, M. (2005), Hedge Funds Portfolio Selection with Higher-order Moments: A Non-parametric Mean-Variance-Skewness-Kurtosis Efficient Frontier, mimeo, 28 pages.

[8] Kohonen, T. (1995), *Self-Organising Maps*, Springer, Berlin.

[9] Merlin, P., Maillet, B. (2005), Completing Hedge Fund Missing Net Asset Values using Kohonen Maps and Constrained Randomization, mimeo Paris-1, 6 pages.

[10] Patton, A. (2004), On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, *Journal of Financial Econometrics*, **vol. 2 (1)**, p. 130-168.

[11] Rousset, P., Maillet, B. (2005), Increasing Reliability of SOMs' Neighbourhood Structure with a Bootstrap Process, mimeo Paris-1, 6 pages.

[12] Samad, T., Harp, S. (1992), Self Organization with Partial Data, *Network*, **vol. 3**, p. 205-212.

[13] Schreiber, T. (1998), Constrained Randomization of Times Series Data, *Physical Review Letter*, **vol. 80 (10)**, p. 2105-2108.

Annexe 3 :

"SOM+EOF for Finding Missing Values"

En collaboration avec Antti Sorjamaa, Amaury Lendasse et Bertrand Maillet.

Paru dans les actes du colloque du *15th European Symposium on Artificial Neural Networks Bruges*, 2007, 115-120.

ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

SOM+EOF for Finding Missing Values

Antti Sorjamaa[1], Paul Merlin[2], Bertrand Maillet[2] and Amaury Lendasse[1]

1- Helsinki University of Technology - CIS
P.O. Box 5400, 02015 HUT - Finland

2- Variances and Paris-1 University CES/CNRS - A.A.Advisors-QCG
106 bv de l'hôpital F-75647 Paris cedex 13 - France

Abstract. In this paper, a new method for the determination of missing values in temporal databases is presented. This new method is based on two projection methods: a nonlinear one (Self-Organized Maps) and a linear one (Empirical Orthogonal Functions). The global methodology that is presented combines the advantages of both methods to get accurate candidates for missing values. An application of the determination of missing values for fund return database is presented.

1 Introduction

The presence of missing values in the underlying time series is a recurrent problem when dealing with databases. Number of methods have been developed to solve the problem and fill the missing values. The methods can be classified into two distinct categories: deterministic methods and stochastic methods.

Self-Organizing Maps [1] (SOM) aim to ideally group homogeneous individuals, highlighting a neighborhood structure between classes in a chosen lattice. The SOM algorithm is based on unsupervised learning principle where the training is entirely stochastic, data-driven. The SOM algorithm allows projection of high-dimensional data to a low-dimensional grid. Through this projection and focusing on its property of topology preservation, SOM allows nonlinear interpolation for missing values.

Empirical Orthogonal Functions (EOF) [2] are deterministic, enabling linear projection to a high-dimensional space. They have also been used to develop models for finding missing data [3]. Moreover, EOF models allow continuous interpolation of missing values, but are sensitive to the initialization.

This paper presents a new methodology, which combines the advantages of both the SOM and the EOF. The nonlinearity property of the SOM is used as a denoising tool and then the continuity property of the EOF method is used to efficiently recover missing data.

2 Self-Organizing Map

The SOM algorithm is based on an unsupervised learning principle, where training is entirely data-driven and no information about the input data is required [1]. Here we use a 2-dimensional network, compound in c units (or code vectors) shaped as a square *lattice*. Each unit of a network has as many weights as the length T of the learning data samples, $\mathbf{x}_n$, $n = 1, 2, ..., N$. All units of a network

can be collected to a weight matrix $\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_c(t)]$ where $\mathbf{m}_i(t)$ is the T-dimensional weight vector of the unit i at time t and t represents the steps of the learning process. Each unit is connected to its neighboring units through neighborhood function $\lambda(\mathbf{m}_i, \mathbf{m}_j, t)$, which defines the shape and the size of the neighborhood at time t. Neighborhood can be constant through the entire learning process or it can change in the course of learning.

Learning starts by initializing the network node weights randomly. Then, for randomly selected sample $\mathbf{x}_{t+1}$, we calculate a Best Matching Unit (BMU), which is the neuron whose weights are closest to the sample. BMU calculation is defined as $\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg\min_{\mathbf{m}_i, i \in I} \{\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|\}$, where $I = [1, 2, ..., c]$ is the set of network node indices, BMU denotes the index of the best matching node and $\|.\|$ is standard Euclidean norm.

If the randomly selected sample includes missing values, the BMU cannot be solved outright. Instead, an adapted SOM algorithm, proposed by Cottrell and Letrémy [4], is used. The randomly drawn sample $\mathbf{x_{t+1}}$ having missing value(s) is split into two subsets $\mathbf{x}_{t+1}^T = NM_{\mathbf{x_{t+1}}} \cup M_{\mathbf{x_{t+1}}}$, where $NM_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are not missing and $M_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are missing. We define a norm on the subset $NM_{\mathbf{x_{t+1}}}$ as

$$\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} = \sum_{k \in NM_{\mathbf{x_{t+1}}}} (\mathbf{x}_{t+1,k} - \mathbf{m}_{i,k}(t))^2, \quad (1)$$

where $\mathbf{x}_{t+1,k}$ for $k = [1, ..., T]$ denotes the k^{th} value of the chosen vector and $\mathbf{m}_{i,k}(t)$ for $k = [1, ..., T]$ and for $i = [1, ..., c]$ is the k^{th} value of the i^{th} code vector.

Then the BMU is calculated with

$$\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg \min_{\mathbf{m}_i, i \in I} \left\{ \|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} \right\}. \quad (2)$$

When the BMU is found the network weights are updated as

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon(t)\lambda\left(\mathbf{m}_{BMU(\mathbf{x}_{t+1})}, \mathbf{m}_i, t\right)[\mathbf{m}_i(t) - \mathbf{x_{t+1}}], \forall i \in I, \quad (3)$$

where $\varepsilon(t)$ is the adaptation gain parameter, which is $]0, 1[$-valued, decreasing gradually with time. The number of neurons taken into account during the weight update depends on the neighborhood function $\lambda(\mathbf{m}_i, \mathbf{m}_j, t)$. The number of neurons, which need the weight update, usually decreases with time.

After the weight update the next sample is randomly drawn from the data matrix and the procedure started again by finding the BMU of the sample. The recursive learning procedure is stopped when the SOM algorithm has converged.

Once the SOM algorithm has converged, we obtain some clusters containing our data. Cottrell and Letrémy proposed to fill the missing values of the dataset by the coordinates of the code vectors of each BMU as natural first candidates for missing value completion:

$$\pi_{(M_{\mathbf{x}})}(\mathbf{x}) = \pi_{(M_{\mathbf{x}})}\left(\mathbf{m}_{BMU(\mathbf{x})}\right), \quad (4)$$

where $\pi_{(M_{\mathbf{x}})}(.)$ replaces the missing values $M_{\mathbf{x}}$ of sample $\mathbf{x}$ with the corresponding values of the BMU of the sample. The replacement is done for every data sample and then the SOM has finished filling the missing values in the data.

3 Empirical Orthogonal Functions

This section presents Empirical Orthogonal Functions (EOF) [2, 5]. In this paper, EOF are used as a denoising tool and for finding the missing values at the same time [3].

The EOF are calculated using standard and well-known Singular Value Decomposition (SVD), $\mathbf{X} = \mathbf{UDV}^* = \sum_{k=1}^{K} \rho_k \mathbf{u}_k \mathbf{v}_k$, where $\mathbf{X}$ is 2-dimensional data matrix, $\mathbf{U}$ and $\mathbf{V}$ are collections of singular vectors $\mathbf{u}$ and $\mathbf{v}$ in each dimension respectively, $\mathbf{D}$ is a diagonal matrix with the singular values ρ in its diagonal and K is the smaller dimension of $\mathbf{X}$ (or the number of nonzero singular values if $\mathbf{X}$ is not full rank). The singular values and the respective vectors are sorted to decreasing order.

When EOF are used to denoise the data, not all singular values and vectors are used to reconstruct the data matrix. Instead, it is assumed that the vectors corresponding to larger singular values contain more data with respect to the noise than the ones corresponding to smaller values [2]. Therefore, it is logical to select q largest singular values and the corresponding vectors and reconstruct the denoised data matrix using only them.

In the case where $q < K$, the reconstructed data matrix is obviously not the same than the original one. The larger q is selected, the more original data, which also includes more noise, is preserved. The optimal q is selected using validation methods, for example [6].

EOF (or SVD) cannot be directly used with databases including missing values. The missing values must be replaced by some initial values in order to use the EOF. This replacement can be for example the mean value of the whole data matrix $\mathbf{X}$ or the mean in one direction, row wise or column wise. The latter approach is more logical when the data matrix has some temporal or spatial structure in its columns or rows.

After the initial value replacement the EOF process begins by performing the SVD and the selected q singular values and vectors are used to build the reconstruction. In order not to lose **any** information, only the missing values of $\mathbf{X}$ are replaced with the values from the reconstruction. After the replacement, the new data matrix is again broken down to singular values and vectors with the SVD and reconstructed again. The procedure is repeated until convergence criterion is fulfilled.

4 Global Methodology

The two methodologies presented in the previous two sections are combined and the global methodology is presented. The SOM algorithm for missing values is first ran through performing a nonlinear projection for finding the missing

values. Then, the result of the SOM estimation is used as initialization for the EOF method.

For The SOM we must select the optimal grid size c and for the EOF the optimal number of singular values and vectors q to be used. This is done using validation, using the same validation set for all combinations of the parameters c and q. Finally, the combination of SOM and EOF that gives the smallest validation error is used to perform the final filling of the data.

Even the SOM as well as the EOF are able to fill the missing values alone, the experimental results demonstrate that together the accuracy is better. The fact that these two algorithms suit well together is not surprising. Two approaches can be considered to understand the complementarity of the algorithms.

Firstly, SOM algorithm allows nonlinear projection. In this sense, even for dataset with complex and nonlinear structure, the SOM code vectors will succeed to capture the nonlinear characteristics of the inputs. However, the projection is done on a low-dimensional grid (in our case two-dimensional) with the possibility of losing the intrinsic information of the data.

EOF method is based on a linear transformation using the Singular Value Decomposition. Because of the linearity of the EOF approach, it will fail to reflect the nonlinear structures of the dataset, but the projection space can be as high as the dimension of the input data and remains continuous.

5 Experimental Results

For illustration, we use a dataset of North American fund returns[1] composed with 679 funds on a 4-year period of 219 weekly values. This gives us a dataset $\mathbf{X}$ of the size 219×679 with a total of 148 701 values. The fund return correspond to the yield of asset values between two consecutive dates as $r_t = \frac{v_{t+1}}{v_t} - 1$, where v_t is the value of the considered asset at time t.

Figure 1 shows 10 rescaled fund values $\left(v_t^{'} = 100 \prod_{i=1}^{t} (1 + r_i)\right)$. The fund values are correlated time series including first order trends. There are no missing values contained in the database.

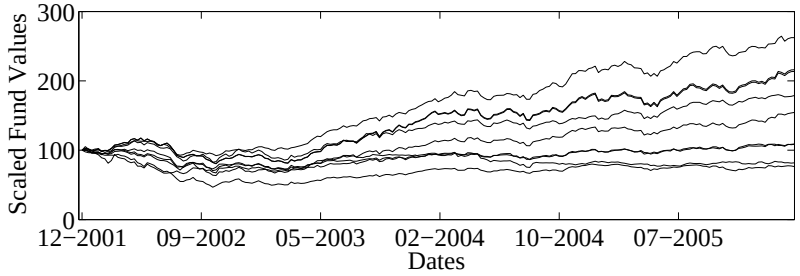


Fig. 1: Rescaled asset values of 10 funds present in the database.

[1]Data provided by Lipper, A Reuters Company.

Before running any experiments, we randomly remove 7.5 percent of the data to a test set. The test set contains 11 152 values. For validation, the same amount of data is removed from the dataset. Therefore, for the model selection and learning we have a database with total of 15 percent missing values.

The Monte Carlo Cross-Validation with 10 folds is used to select the optimal parameters for the SOM, the EOF and the SOM+EOF method. The 10 selected validation sets are the same for each method. All validation errors are shown in Figure 2. In the case of the SOM+EOF, the errors shown are minimum errors after EOF with different SOM sizes.

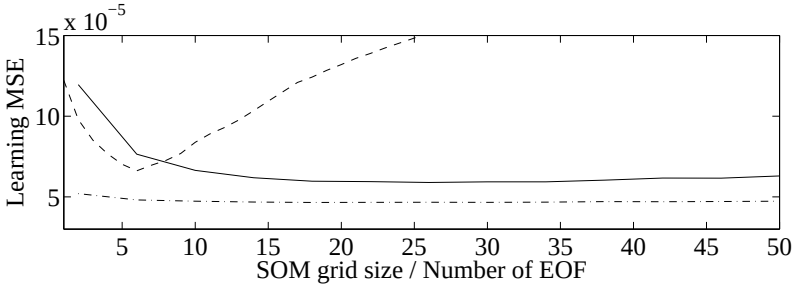


Fig. 2: Validation errors w.r.t. SOM size or number of EOF. SOM validation error (Solid line), EOF (dashed line) and SOM+EOF (dash-dotted line).

The optimal size of the SOM grid is found to be 28×28, which is a total of 784 units. Therefore, we have more code vectors in the SOM than observations (629). It means that we have nonlinear interpolation between observations and better approximation of the missing values with more units than data.

When the EOF is performed alone, initial values are substituted as the column means of the original matrix, calculated only with the known values. From the Figure 2 the smallest error with the EOF method is achieved with q equal to 6. This number of EOF is very small compared to the maximum of 219 EOF, which is the smaller dimension of the data. It suggests quite strong noise influence in the data and that there is only a small number of efficient EOF needed to represent the denoised data.

The smallest error achieved with the SOM+EOF method is with SOM grid size 27×27 and with EOF parameter q equal to 39. The number of selected EOF is much larger with the SOM initialization than with the column mean initialization. It suggests there are more efficient EOF to use in the approximation of the missing values than with the plain column mean initialization and that the SOM has already denoised the data. The optimal SOM grid size in the SOM+EOF method is found out to be roughly the same size than when performing the SOM alone. It is quite intuitive to think that the best possible filling achieved with SOM is enhanced with linear, high-dimensional projection of the EOF. From the Figure 2 it is clearly notable that with every SOM size the SOM+EOF method gives lower validation error than either SOM or EOF alone.

Table 1 contains the validation and test errors of all three methods.

Table 1: Learning and Test Errors for SOM, EOF and SOM+EOF.

10^{-5}	Learning Error	Test Error
SOM	5.83	5.57
EOF	6.61	6.13
SOM + EOF	4.63	4.34

From the Table 1, we can see that the SOM+EOF outperforms the SOM reducing the validation error by 21 percent and the test error by 22 percent. The EOF alone is not performing as well as the SOM alone.

6 Conclusion

In this paper, we have compared 3 methods for finding missing values in temporal databases. The methods are Self-Organizing Maps (SOM), Empirical Orthogonal Function (EOF) and the combination of the two SOM+EOF.

The advantages of the SOM include the ability to perform nonlinear projection of high-dimensional data to lower dimension with interpolation between discrete data points. For the EOF, the advantages include high-dimensional linear projection of high-dimensional data and the speed and the simplicity of the method. The SOM+EOF includes the advantages of both individual methods, leading to a new accurate approximation methodology for finding the missing values. The performance obtained in test show the accuracy of the new methodology.

It has also been shown experimentally that the optimal number of code vectors used in the SOM has to be larger than the number of observations. It is necessary in order to take the advantage of the self-organizing property of the SOM and the interpolation ability for finding the missing data.

For further work, the modifications and performance upgrades for the global methodology are fine-tuned for different types of datasets. The methodology will then be applied to datasets from climatology.

References

[1] Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 1995.

[2] R. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, 1988.

[3] J. Boyd, E. Kennelly, and P. Pistek. Estimation of eof expansion coefficients from incomplete data. *Deep Sea Research*.

[4] Marie Cottrell and Patrick Letrémy. Missing values: Processing with the kohonen algorithm. pages 489–496. Applied Stochastic Models and Data Analysis, Brest, France, 17-20 May, 2005.

[5] SOM+EOF Toolbox: http://www.cis.hut.fi/projects/tsp/?page=Downloads.

[6] Amaury Lendasse, V. Wertz, and Michel Verleysen. Model selection with cross-validations and bootstraps - application to time series prediction with rbfn models. In *LNCS*, number 2714, pages 573–580, Berlin, 2003. ICANN/ICONIP (2003), Springer-Verlag.

Annexe 4 :

*"Hedge Funds Portfolio Selection with Higher-order Moments: A Non-parametric Mean-Variance-Skewness-*Kurtosis *Efficient Frontier"*

En collaboration avec Emmanuel Jurczenko et Bertrand Maillet.

Paru dans *Multi-moment Asset Allocation and Pricing Models*, Jurczenko-Maillet (Eds), John Wiley & Sons New-York, 2006, Chapitre 3, 51-66.

3

Hedge Fund Portfolio Selection with Higher-order Moments: A Nonparametric Mean–Variance–Skewness–*Kurtosis* Efficient Frontier

Emmanuel Jurczenko, Bertrand Maillet and Paul Merlin

ABSTRACT

This chapter proposes a nonparametric optimisation criterion for the static portfolio selection problem in the mean–variance–skewness–*kurtosis* space. Following the work of Briec *et al.* (2004 and 2006), a shortage function is defined, in the four-moment space, that looks simultaneously for improvements in the expected portfolio return, variance, skewness and *kurtosis* directions. This new approach allows us to optimise multiple competing and often conflicting asset allocation objectives within a mean–variance–skewness–*kurtosis* framework. The global optimality is here guaranteed for the resulting optimal portfolios. We also establish a link to a proper indirect four-moment utility function. An empirical application on funds of hedge funds serves to show a three-dimensional representation of the primal nonconvex mean–variance–skewness–*kurtosis* efficient portfolio set and to illustrate the computational tractabilty of the approach.

3.1 INTRODUCTION

The mean–variance decision criterion proposed by Markowitz (1952) is inadequate for allocating wealth when dealing with hedge funds. Not only are hedge fund return distributions asymmetric and leptokurtic, but they also display significant coskewness and co*kurtosis* with the returns of other asset classes, due to the option-like features of alternative investments (see Weisman, 2002; Goetzmann *et al.*, 2004; Agarwal and Naik, 2004 and Davies *et al.*, 2004).

Different approaches have been developed in the financial literature to incorporate the individual preferences for higher-order moments into optimal asset allocation problems, though no single conclusive approach seems to have emerged yet. These approaches can be divided between primal and dual program for determining the mean–variance–skewness–*kurtosis* efficient frontier.

Multi-moment Asset Allocation and Pricing Models Edited by E. Jurczenko and B. Maillet

Davies *et al.* (2005) and Berényi (2001 and 2002) use polynomial goal programming (PGP) to determine the set of the mean–variance–skewness–*kurtosis* efficient funds of hedge funds.[1] A shortcoming of this primal approach is that the allocation problem solved in the PGP cannot be related precisely to the expected utility function. In particular, the choice of the parameters used to weight the moment deviations is not related to the parameters of the utility function. Another drawback of the estimation of the four-moment efficient frontier via multi-objective programming is that it is not compliant with the Pareto-optimal definition of an efficient portfolio frontier. Indeed, minimising deviations from the first four moments simultaneously only guarantees a solution close to the mean–variance–skewness–*kurtosis* efficient frontier. Consequently, no portfolio performance measure can be inferred from the exercise. Some primal contributions solve analytically the mean–variance–skewness–*kurtosis* portfolio optimisation problem. For example, Athayde and Flôres (2002), Adcock (2003) and Jurczenko and Maillet in Chapter 6 look for the analytical solution characterising the minimum variance frontier in the mean–variance–skewness–*kurtosis* space, assuming shorting, with the objective of minimising the variance for a given mean, skewness and *kurtosis*. These approaches are, however, partial since they focus mainly on one objective of the mean–variance–skewness–*kurtosis* optimisation program at the cost of the others.[2]

Dual approaches start instead from a particular specification of the indirect mean–variance–skewness–*kurtosis* utility by using a Taylor series expansion of the investors' objective functions to determine the optimal portfolios (see, for instance, Guidolin and Timmermann, 2005; Jondeau and Rockinger, 2003 and 2006 and Jurczenko and Maillet in Chapter 1 of this book.[3] While such approaches have been used extensively in empirical applications to test multi-moment CAPM, they suffer from severe limitations in the context of hedge fund asset allocations. The Taylor series expansion may converge to the expected utility under restrictive conditions only. For some utility functions (such as the exponential one), the expansion converges for all possible levels of return, while for others (e.g. logarithm-power type utility functions), convergence is ensured only over a restricted range that may be problematic for some alternative investments due to the presence of leverage effects. In addition, the truncation of the Taylor series raises several difficulties. In particular, there is generally no rule for selecting the order of truncation. The inclusion of an additional moment does not necessarily improve the quality of the approximation (see Chapter 1). Dual approaches are also hampered by the lack of knowledge of the individual preferences for the first four moments of the portfolio return distribution and suffer from their lack of integration with the primal approaches briefly outlined above. Moreover, since the mean–variance–skewness–*kurtosis* efficient frontier is a nonconvex surface, previous parametric primal and dual approaches can only guarantee local optimal solutions to the portfolio optimisation problems in the four-moment space, not a global one. They inevitably require one to convexify some part of the nonconvex four-moment efficient frontier by using *ad hoc* moment restrictions, separating return distributions or separating utility functions (see Rubinstein, 1973; Ingersoll, 1987 and Athayde and Flôres, 2004). Dual approaches carry, in particular, the risk that certain target portfolios based upon particular specifications of the utility function are infeasible in practice. As the dimensionality of the portfolio

[1] For studies of the use of this approach in the mean–variance–skewness portfolio selection case, see Lai (1991), Chunhachinda *et al.* (1997), Wang and Xia (2002), Chang *et al.* (2003) and Sun and Yan (2003).

[2] See also Simaan (1993), Gamba and Rossi (1997, 1998a and 1998b), Pressacco and Stucchi (2000) and Jurczenko and Maillet (2001), for similar optimisation programmes in the mean–variance–skewness space.

[3] See Harvey *et al.* (2004) for the mean–variance–skewness portfolio selection case.

selection problem increases, it then becomes difficult to develop a geometric interpretation of the portfolio efficient frontier and to select the preferred portfolio among the boundary points.

To circumvent these problems, we use a particular distance function – the shortage function – to incorporate investors' preferences for higher moments into the optimal construction of a fund of hedge funds. The shortage function enables us to solve for the multiple conflicting and competing allocation objectives without assuming a detailed knowledge of the preference parameters of the indirect utility function. It integrates the primal and the dual approaches.

The shortage function, first introduced by Luenberger (1995) in production theory, is a distance function that looks simultaneously for reduction in inputs and expansion in outputs, and that is dual to the profit function. It offers a perfect representation of multidimensional choice sets and can position any point relative to the boundary frontier of the choice set. It has been used subsequently by Morey and Morey (1999) and Briec *et al.* (2004) for gauging the performance of funds in the mean–variance framework, and more recently by Briec *et al.* (2006) for solving portfolio selection problems involving significant degrees of skewness. In this chapter, we extend the shortage function from the mean–variance–skewness space to the mean–variance–skewness–*kurtosis* one to take into account the aversion to *kurtosis* in addition to individual preferences for expected return, variance and skewness. The shortage function projects any (in)efficient portfolio exactly onto the four-dimensional mean–variance–skewness–*kurtosis* portfolio frontier. It rates portfolio performance by measuring a distance between a portfolio and its optimal projection onto the primal mean–variance–skewness–*kurtosis* efficient frontier. Following the same line of reasoning as Briec *et al.* (2004 and 2006), we prove that our shortage function achieves a global optimum on the boundary of the nonconvex mean–variance–skewness–*kurtosis* portfolio frontier and establish a duality result between the shortage function and the indirect mean–variance–skewness–*kurtosis* utility function.

Thanks to the global optimality and duality results, the shortage function approach stands out compared to the existing four-moment primal and dual approaches, which only guarantee a local optimal solution to the investor's portfolio optimisation programme. Moreover, our multi-moment portfolio selection approach is more general than the previous ones since we are not assuming the existence of a riskless asset and forbidding short-sales.

The remainder of the chapter is organised as follows. In Section 3.2 we describe the optimal hedge fund portfolio selection program within a four-moment framework. In Section 3.3 we introduce the shortage function, study its axiomatic properties and establish the link between the shortage function and the indirect mean–variance–skewness–*kurtosis* utility function. Section 3.4 describes the data and hedge fund classification and provides illustrative empirical results. Section 3.5 concludes. Proofs are presented separately in the appendix.

3.2 PORTFOLIO SELECTION WITH HIGHER-ORDER MOMENTS

We consider the problem of an investor selecting a portfolio from N risky assets (with $N \geq 4$) in the mean–variance–skewness–*kurtosis* framework (see Chapter 1). We assume that the investor does not have access to a riskless asset, implying that the portfolio weights must sum to one. In addition, we impose a no short-sale portfolio constraint: asset positions

must be non-negative. Let $\mathbf{w}_p$ and $\mathbf{E}$ denote respectively the $(N \times 1)$ vector of weights and of expected returns for the N risky assets in the portfolio p; Ω the nonsingular $(N \times N)$ variance–covariance matrix of the risky assets; and Σ and Γ represent respectively the $\left(N \times N^2\right)$ skewness–coskewness matrix and the $\left(N \times N^3\right)$ *kurtosis–cokurtosis* matrix of the N risky asset returns, defined as (Athayde and Flôres, 2004 and Chapter 2 of this book):

$$\begin{cases} \underset{(N\times N^2)}{\sum} = (\Sigma_1 \Sigma_2 \cdots \Sigma_N) \\ \underset{(N\times N^3)}{\Gamma} = (\Gamma_{11}\Gamma_{12}\cdots\Gamma_{1N}|\,\Gamma_{21}\Gamma_{22}\cdots\Gamma_{2N}\,|\ldots|\,\Gamma_{N1}\Gamma_{12}\cdots\Gamma_{NN}) \end{cases} \tag{3.1}$$

where Σ_k and Γ_{kl} are the $(N \times N)$ associated submatrices of Σ and Γ, with elements $\left(s_{ijk}\right)$ and $\left(\kappa_{ijkl}\right)$, with $(i, j, k, l) \in (IN^*)^4$, and the sign $\otimes$ stands for the Kronecker product.[4]

It should be noted that, because of the symmetries, not all the elements of these matrices need to be computed. Only $N(N+1)/2$ elements of the $(N \times N)$ variance–covariance matrix must be computed. Similarly the skewness–coskewness and *kurtosis*–co*kurtosis* matrices have dimensions $\left(N \times N^2\right)$ and $\left(N \times N^3\right)$, but only $N(N+1)(N+2)/6$ and $N(N+1)(N+2)(N+3)/24$ elements are independent.[5]

The set of the feasible portfolios $\Im_p$ can be expressed as follows:

$$\Im_p = \left\{\mathbf{w}_p \in IR^N : \mathbf{w}_p'\mathbf{1} = \mathbf{1} \text{ and } \mathbf{w}_p \geq \mathbf{0}\right\} \tag{3.2}$$

where $\mathbf{w}_p'$ is the $(\mathbf{1} \times N)$ transposed vector of the investor's holdings of risky assets and $\mathbf{1}$ is the $(N \times \mathbf{1})$ unitary vector.

The mean, variance, skewness and *kurtosis* of the return of a given portfolio p belonging to $\Im_p$ are respectively given by:

$$\begin{cases} E\left(R_p\right) = E\left[\sum_{i=1}^{N}\left(w_{pi}\,R_i\right)\right] = \mathbf{w}_p'\,\mathbf{E} \\ \sigma^2\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^2\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{pi}w_{pj}\,\sigma_{ij} = \mathbf{w}_p'\Omega\mathbf{w}_p \\ s^3\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^3\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} w_{pi}w_{pj}w_{pk}s_{ijk} = \mathbf{w}_p'\,\Sigma\left(\mathbf{w}_p \otimes \mathbf{w}_p\right) \\ \kappa^4\left(R_p\right) = E\left\{\left[R_p - E\left(R_p\right)\right]^4\right\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} w_{pi}w_{pj}w_{pk}w_{pl}\kappa_{ijkl} \\ \qquad\qquad = \mathbf{w}_p'\,\Gamma\left(\mathbf{w}_p \otimes \mathbf{w}_p \otimes \mathbf{w}_p\right) \end{cases} \tag{3.3}$$

[4] Let $\mathbf{A}$ be an $(n \times p)$ matrix and $\mathbf{B}$ an $(m \times q)$ matrix. The $(mn \times pq)$ matrix $\mathbf{A} \otimes \mathbf{B}$ is called the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}\mathbf{B} & a_{N2}\mathbf{B} & \cdots & a_{NN}\mathbf{B} \end{pmatrix}$$

where the sign $\otimes$ stands for the Kronecker product.

[5] For $N = 4$, where these matrices have respectively 16, 64 and 256 terms, ten different elements for the variance–covariance matrix, 20 elements for the skewness–coskewness matrix and 35 elements for the *kurtosis–cokurtosis* matrix are to be computed.

with, $\forall(i, j, k, l) \in [1, \ldots, N]^4$:

$$\begin{cases} R_p = \sum_{i=1}^{N} w_{pi} R_i \\ \sigma_{ij} = E\left\{[R_i - E(R_i)]\left[R_j - E\left(R_j\right)\right]\right\} \\ s_{ijk} = E\left\{[R_i - E(R_i)]\left[R_j - E\left(R_j\right)\right][R_k - E(R_k)]\right\} \\ \kappa_{ijkl} = E\left\{[R_i - E(R_i)]\left[R_j - E\left(R_j\right)\right][R_k - E(R_k)][R_l - E(R_l)]\right\} \end{cases}$$

where $\left(w_{pi}\right)$, (R_i), $\left(\sigma_{ij}\right)$, $\left(s_{ijk}\right)$ and $\left(\kappa_{ijkl}\right)$ represent, respectively, the weight of the asset i in the portfolio p, the return on the asset i, the covariance between the returns of asset i and j, the coskewness between the returns of asset i, j and k and the *cokurtosis* between the returns of asset i, j, k and l, with $(i \times j \times k \times l) = (IN^*)^4$.

Following Markowitz (1952) leads to the following disposal representation, denoted $\mathfrak{D}_p$, of the set of the feasible portfolios in the mean–variance–skewness–*kurtosis* space (see Briec *et al.,* 2004 and 2006):

$$\mathfrak{D}_p = \left\{\mathbf{m}_p : \mathbf{w}_p \in \mathfrak{F}_p\right\} + \left[IR_+ \times (-IR_+) \times IR_+ \times (-IR_+)\right] \tag{3.4}$$

with:

$$\mathbf{m}_p = \left[\kappa^4\left(R_p\right) s^3\left(R_p\right) \sigma^2\left(R_p\right) E\left(R_p\right)\right]'$$

where $\mathbf{m}_p$ is the (4×1) vector of the first four moments of the portfolio return p. This disposal representation is necessary to ensure the convexity of the feasible portfolio set in the mean–variance–skewness–*kurtosis* space.

The four-moment (weakly) efficient portfolio frontier is then defined as follows:

$$\mathfrak{M}_p = \left\{\mathbf{m}_p : \mathbf{m}_q > \mathbf{m}_p \Rightarrow \mathbf{m}_q \notin \mathfrak{D}_p\right\}$$

The weakly efficient frontier is the set of all the mean, variance, skewness, *kurtosis* quadruplets that are not strictly dominated in the four-dimensional space.

The set of the weakly efficient portfolios in the four-moment case is then given in the simplex as:

$$\mathfrak{E}_p = \left\{\mathbf{w}_p \in \mathfrak{F}_p : \mathbf{m}_p \in \mathfrak{M}_p\right\} \tag{3.5}$$

By analogy with production theory (Luenberger, 1995), the next section introduces the shortage function as an indicator of the mean–variance–skewness–*kurtosis* portfolio (in)efficiency.

3.3 THE SHORTAGE FUNCTION AND THE MEAN–VARIANCE–SKEWNESS–*KURTOSIS* EFFICIENT FRONTIER

In production theory, the shortage function measures the distance between some point of the production possibility set and the efficient production frontier (Luenberger, 1995).

The properties of the set of portfolio return moments on which the shortage function is defined have already been discussed in the mean–variance plane by Briec *et al.* (2004) and in the mean–variance–skewness space by Briec *et al.* (2006). It is now possible to extend their definitions to get a portfolio efficiency indicator in the four-moment case.

The shortage function associated with a feasible portfolio p with reference to the direction vector $\mathbf{g}$ in the mean–variance–skewness–*kurtosis* space is the real-valued function $S_{\mathbf{g}}(.)$ defined as:

$$S_{\mathbf{g}}\left(\mathbf{w}_p\right) = \sup\left\{\delta : \mathbf{m}_p + \delta\mathbf{g} \in \mathfrak{D}_p, \mathbf{g} \in IR_+ \times IR_- \times IR_+ \times IR_-\right\} \tag{3.6}$$

with:

$$\begin{cases} \mathbf{m}_p = \left(\kappa^4\left(R_p\right) s^3\left(R_p\right) \sigma^2\left(R_p\right) E\left(R_p\right)\right)' \\ \mathbf{g} = \left(-g_\kappa + g_s - g_\sigma + g_E\right)' \end{cases}$$

where $\mathbf{g}$ is the directional vector in the four-moment space.

The use of the shortage function in the mean–variance–skewness–*kurtosis* space can only guarantee weak efficiency for a portfolio, since it does not exclude projections on the vertical and horizontal parts of the frontier allowing for additional improvements. Furthermore, portfolios that are weakly dominated in terms of their expected return, variance, skewness and *kurtosis* are only weakly mean–variance–skewness–*kurtosis* efficient.

The disposal representation of the feasible portfolio set can be used to derive the lower bound of the true unknown four-moment efficient frontier through the computation of the associated portfolio shortage function. Let us consider a specific portfolio $\mathbf{w}_k$ from a sample of P portfolios – or assets – $\left(\mathbf{w}_p\right)$, with $p = [1, \ldots, P]$, whose performances need to be evaluated in the four-moment dimensions. The shortage function for this portfolio is then computed by solving the following quartic optimisation program:

$$\begin{aligned} \mathbf{w}_p^* &= \underset{\mathbf{w}_p}{\text{Arg}}\left\{\text{Max}\,\delta\right\} \\ \text{s.t.} &\begin{cases} E\left(R_k\right) + \delta g_E \leq E\left(R_p\right) \\ \sigma^2\left(R_k\right) - \delta\, g_\sigma \geq \sigma^2\left(R_p\right) \\ s^3\left(R_k\right) + \delta\, g_s \leq s^3\left(R_p\right) \\ \kappa^4\left(R_k\right) - \delta\, g'_\kappa \geq \kappa^4\left(R_p\right) \\ \mathbf{w}'_p\,\mathbf{1} = 1 \\ \mathbf{w}_p \geq \mathbf{0} \end{cases} \end{aligned} \tag{3.7}$$

where $\mathbf{w}_{p^*}$ is the $(N \times \mathbf{1})$ efficient portfolio weight vector that maximises the performance, risk, skewness and *kurtosis* relative improvement over the evaluated portfolio in the direction vector $\mathbf{g}$. Using the vectorial notation of the portfolio return higher moments (3.1) and using the first four moments of the evaluated portfolio k in the expression of the

direction vector $\mathbf{g}$, the nonparametric portfolio optimisation program (3.7) can then be restated as:

$$\mathbf{w}_p^* = \underset{\mathbf{w}_p}{\text{Arg}}\,\{\text{Max}\,\delta\}$$
$$\text{s.t.}\begin{cases} E(R_k) + \delta E(R_k) \leq \mathbf{w}_p' \mathbf{E} \\ \sigma^2(R_k) - \delta \sigma^2(R_k) \geq \mathbf{w}_p' \Omega \mathbf{w}_p \\ s^3(R_k) + \delta s^3(R_k) \leq \mathbf{w}_p' \Sigma (\mathbf{w}_p \otimes \mathbf{w}_p) \\ \kappa^4(R_k) - \delta \kappa^4(R_k) \geq \mathbf{w}_p' \Gamma (\mathbf{w}_p \otimes \mathbf{w}_p \otimes \mathbf{w}_p) \\ \mathbf{w}_p' \mathbf{1} = \mathbf{1} \\ \mathbf{w}_p \geq \mathbf{0} \end{cases} \tag{3.8}$$

with:

$$\mathbf{g} = \left[-\kappa^4(R_k)\, s^3(R_k) - \sigma^2(R_k)\, E(R_k)\right]'$$

The optimisation programs (3.7) and (3.8) are special cases of the following standard nonlinear quartic program:

$$\mathbf{z}^* = \underset{\mathbf{z}}{\text{Arg}}\,\{\text{Min}\,\mathbf{c}'\mathbf{z}\}$$
$$\text{s.t.}\begin{cases} L_j(\mathbf{z}) \leq \alpha_j \\ Q_k(\mathbf{z}) \leq \beta_k \\ C_l(\mathbf{z}) \leq \gamma_l \\ Q_q^*(\mathbf{z}) \leq \gamma_q \end{cases} \tag{3.9}$$

where $\mathbf{z} \in IR^p$, $L_j(.)$ is a linear map for $j = [1, \ldots, J]$, $Q_k(.)$ is a positive semi-definite quadratic form for k, $k = [1, \ldots, K]$, $C_l(.)$ is a cubic form for l, $l = [1, \ldots, L]$, and $Q_q^*(.)$ is a quartic form for q, $q = [1, \ldots, Q]$. In the case of the portfolio optimisation programme (3.8), $p = n, J = K = L = Q = 1$. The programme is not a standard convex nonlinear optimisation problem.

Due to the non-convex nature of the optimisation program, we need to state the necessary and sufficient conditions showing that a local optimal solution of (3.8) is also a global optimum (see the appendix).

Despite the nonconvex nature of the mean–variance–skewness–*kurtosis* portfolio selection program, the shortage function maximisation achieves a global optimum for the cubic program. This makes the shortage approach superior to the other primal and dual approaches of the mean–variance–skewness–*kurtosis* efficient set listed in the introduction, since those guarantee only a local optimum solution. To our knowledge, it encompasses also all the existing primal portfolio selection methods with higher-order moments considered in the financial literature. In the next section, we illustrate the shortage function approach in the case of hedge fund selection.

3.4 DATA AND EMPIRICAL RESULTS

Figures 3.1 to 3.8 provide different geometrical representations of a four-moment efficient frontier obtained after the optimisation of hedge funds in 1296 (6^4) directions using our directional distance function approach.

The original data – provided by HFR – consist of monthly net asset values of hedge funds (expressed in EUR) since January 1995. The maximum number of funds in the database is reached in September 2004 (4279 funds were observed). We then delete funds with missing values and normalise fund values to index 100 at the beginning of the final sample. At the end, 20 funds remain – which can be considered a fair number of funds for a fund of hedge funds – and the number of observations considered is 120 (from January 1995 to January 2005) – which can be considered long enough for this kind of application.

Figure 3.1 (3.2) represents the four-moment optimal portfolios in the mean–variance–third moment space (the mean–variance–fourth moment space). The maximisation of the expected return leads, at the optimum, to an increase of the variance – as in the Markowitz case – and the maximisation (minimisation) of the expected return (variance) implies, for hedge funds, also an increase (decrease) in the skewness (fourth moment). That is, while the individual preferences for higher-order moments cause high *kurtosis* and standard deviation

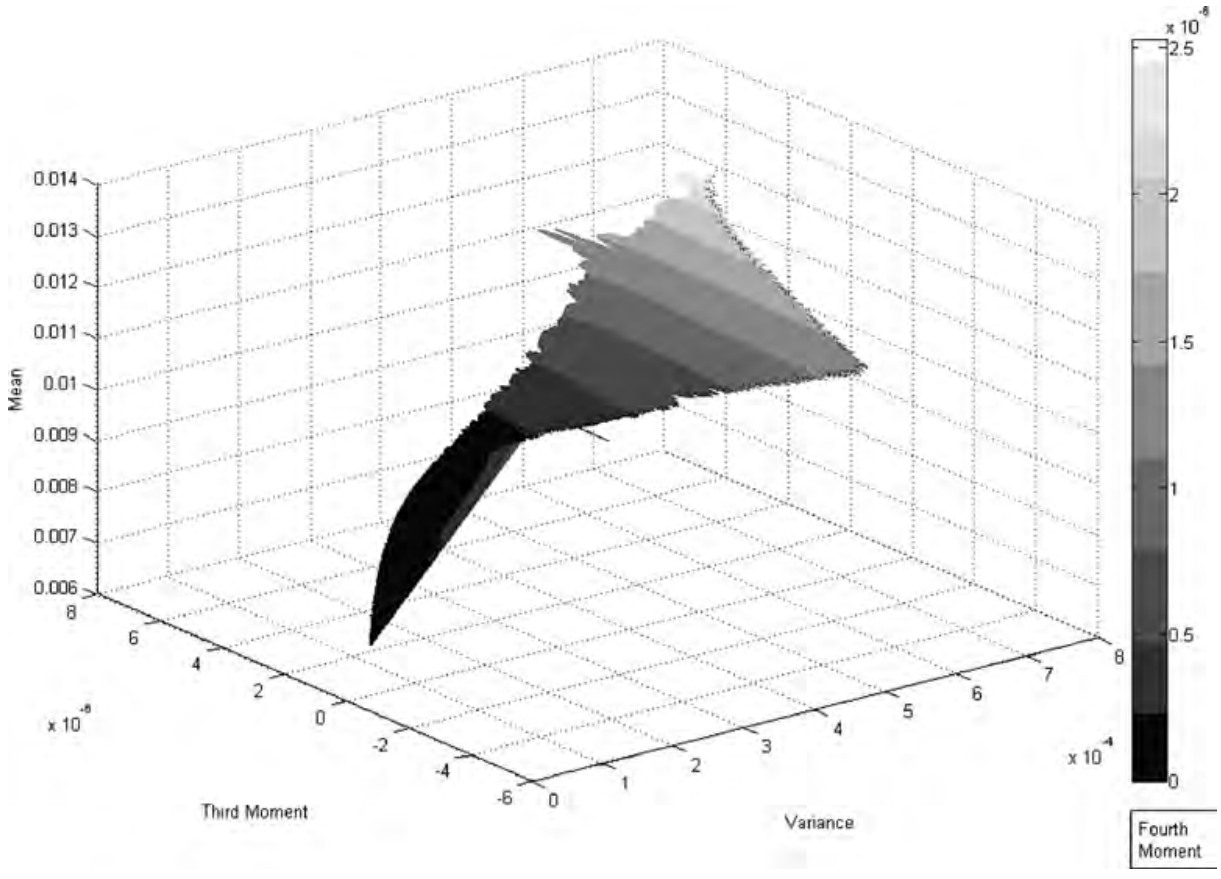


Figure 3.1 Mean–variance–skewness–*kurtosis* constrained efficient frontier in the mean–variance–third moment space. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. The constrained efficient frontier is obtained after optimisation of 20 hedge funds in 1296 directions. Grey shading represents the level of the fourth noncentral moment.

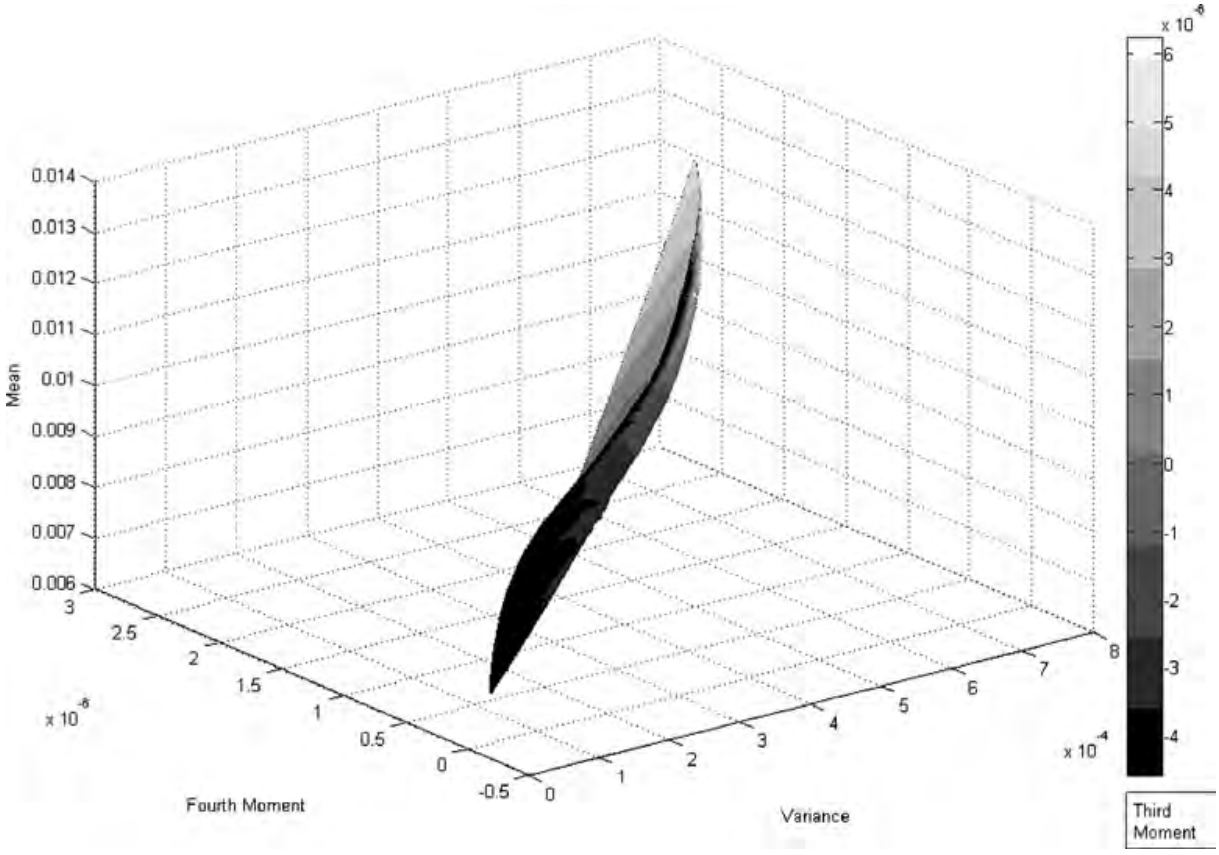


Figure 3.2 Mean–variance–skewness–*kurtosis* constrained efficient frontier in the mean–variance–fourth moment space. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. The constrained efficient frontier is obtained after optimisation of 20 hedge funds in 1296 directions. Grey shading represents the level of the third noncentral moment.

to be traded for higher expected return and skewness, hedge fund returns do not seem to exhibit the same type of trade-offs between even or odd moments that are typically observed in the underlying securities markets. These results are confirmed by Figures 3.3 to 3.8, which present the coordinates of the mean–variance–skewness–*kurtosis* efficient portfolios in several moment planes. Indeed, Figure 3.3 shows that mean–variance efficient portfolios are efficient in terms of *kurtosis*, but not necessarily in terms of skewness. For instance, given the mean, it is possible to increase the skewness at the cost of the variance. It is, however, not possible to decrease the fourth-order moment when controlling for variance. Likewise, for intermediate or extreme levels of variance, it is possible to increase the skewness of an optimal portfolio at the cost of its expected return. These observations contradict the point raised by Davies *et al.* (2004 and 2005) and Andersen and Sornette (2001), namely that mean–variance optimisers may be nothing more than skewness minimisers and *kurtosis* maximisers. Figures 3.4 and 3.5 document the existence in the four-moment efficient set of a V-shaped relationship between the third moment and the mean and the variance, and Figures 3.6 and 3.7 illustrate the existence of a concave and positive relation between the optimal fourth moment and the expected portfolio return and variance.[6]

[6] Since the properties of the efficient set depend heavily on the technological characteristics of the underlying assets, further investigations on hedge fund strategies and asset classes are required to assess the generality of our empirical findings.

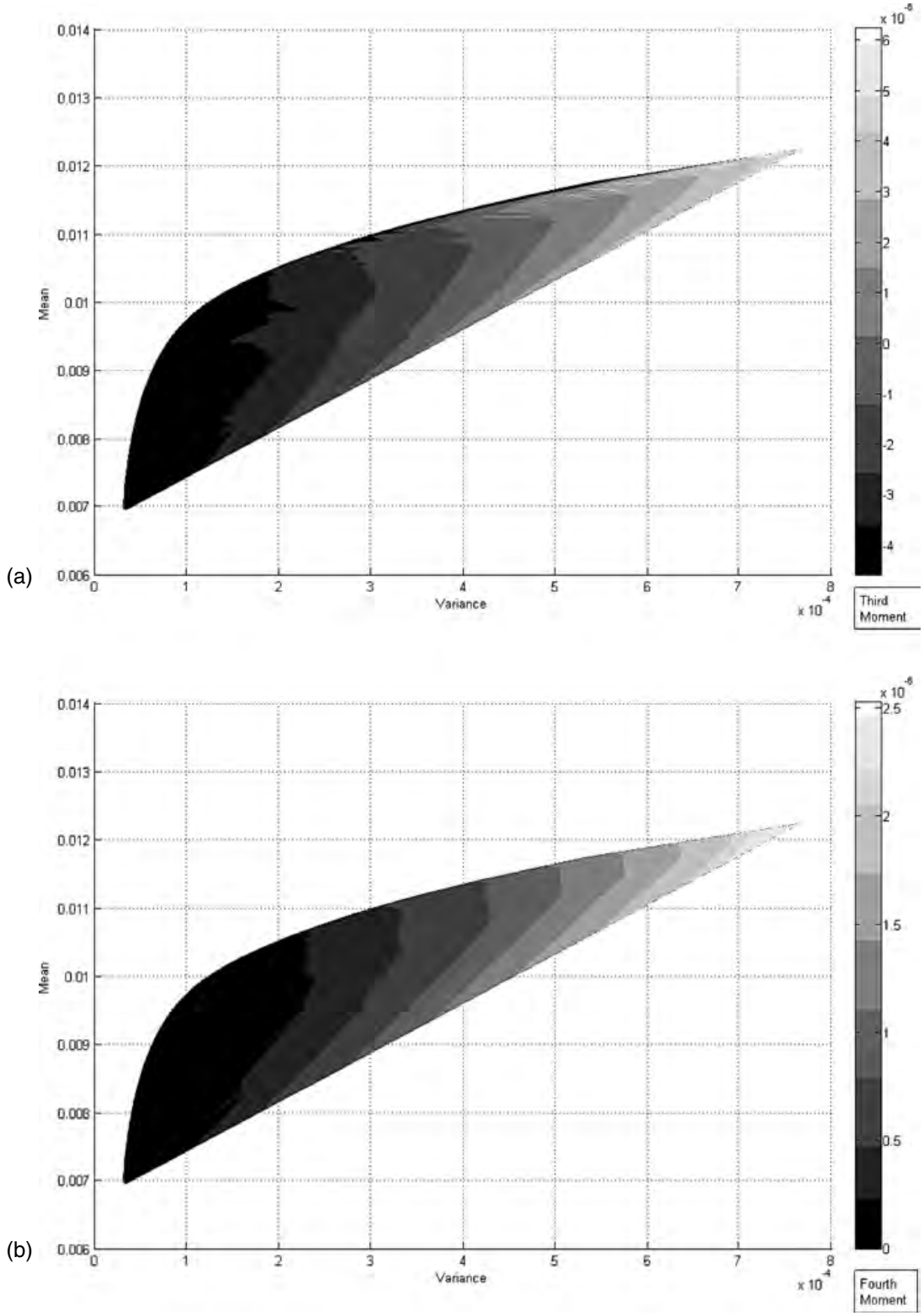


Figure 3.3 Mean–variance–skewness–*kurtosis* constrained efficient frontier in the mean–variance plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. The constrained efficient frontier is obtained after optimisation of 20 hedge funds in 1296 directions. (a) grey shading represents the level of the third noncentral moment; (b) grey shading represents the level of the fourth noncentral moment.

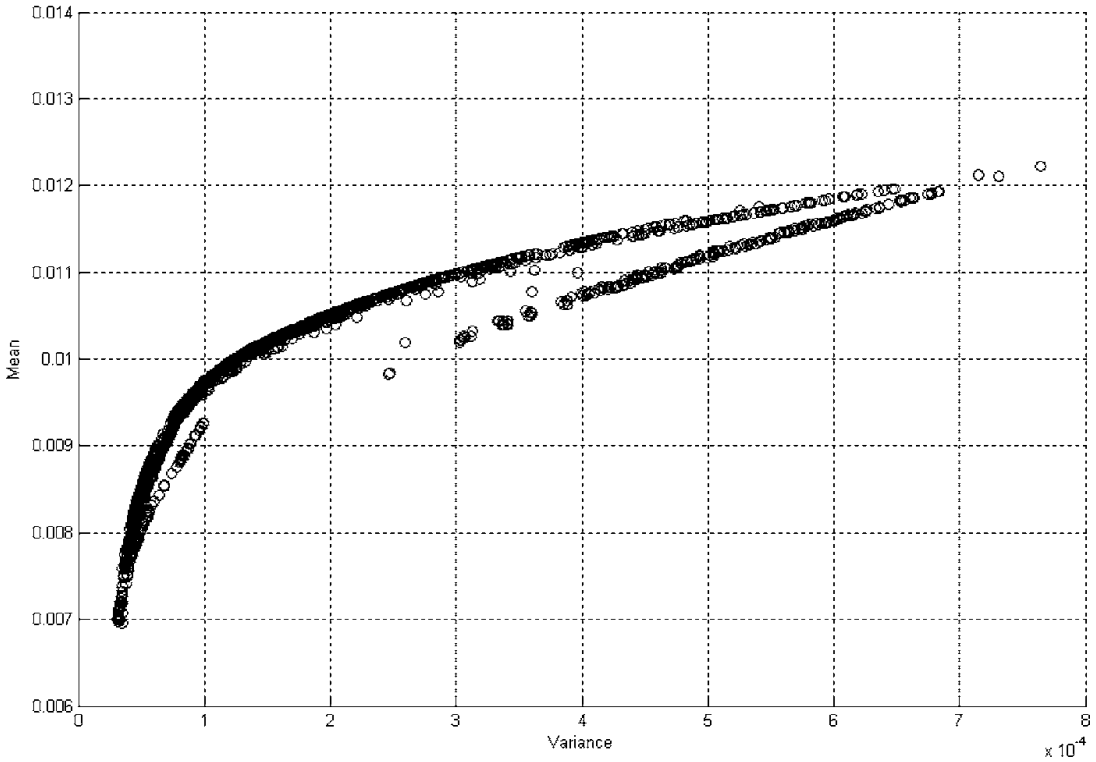


Figure 3.4 Mean–variance–skewness–*kurtosis* constrained efficient portfolios in the mean–variance plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. Optimal points are obtained after optimisation of 20 hedge funds in 1296 directions.

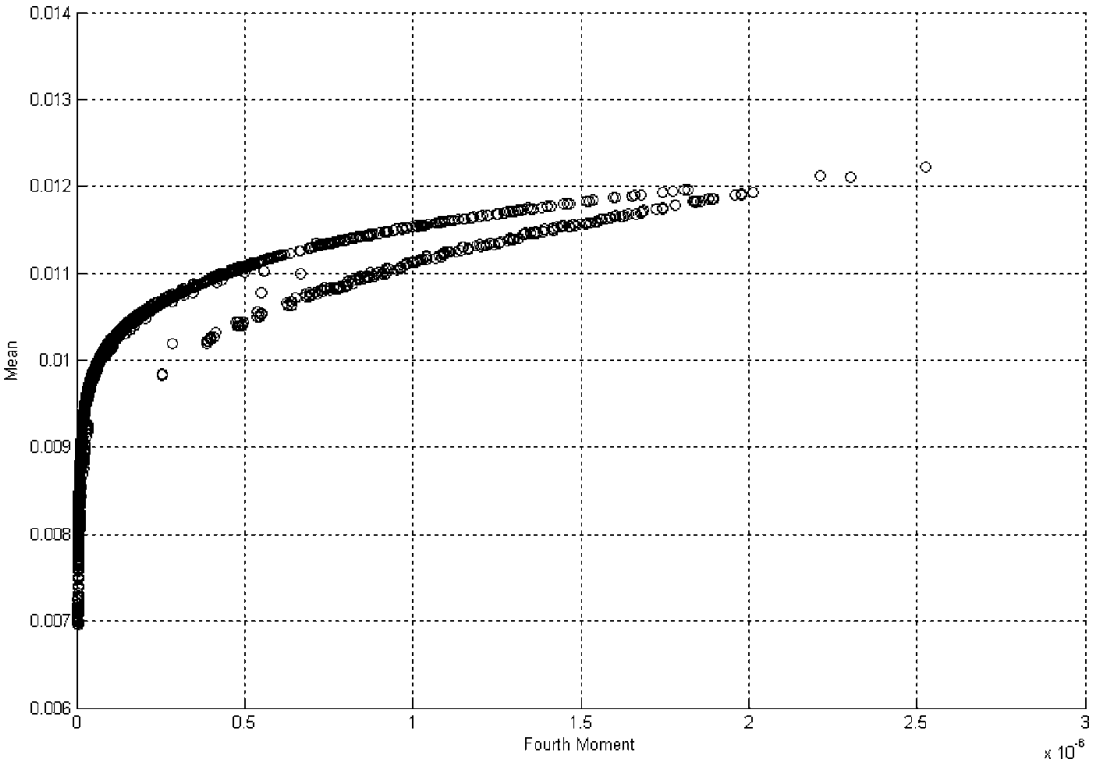


Figure 3.5 Mean–variance–skewness–*kurtosis* efficient portfolios in the mean–fourth moment plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. Optimal points are obtained after optimisation of 20 hedge funds in 1296 directions.

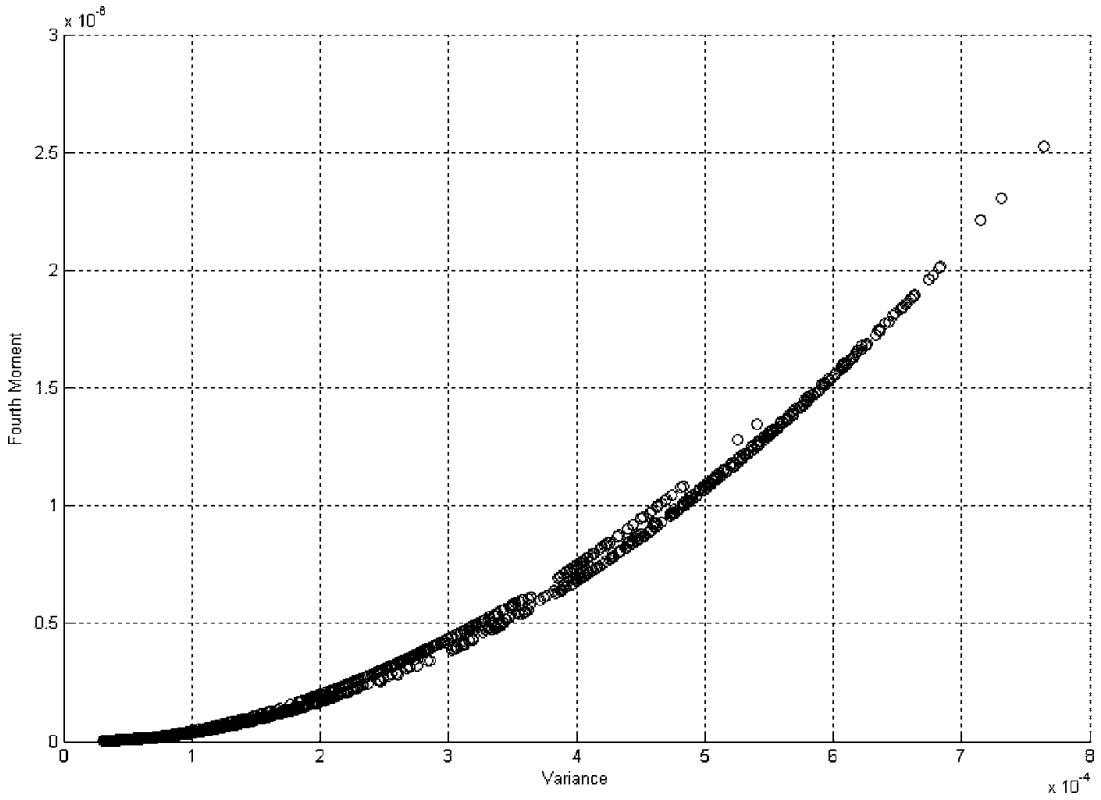


Figure 3.6 Mean–variance–skewness–*kurtosis* efficient portfolios in the variance–fourth moment plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. Optimal points are obtained after optimisation of 20 hedge funds in 1296 directions.

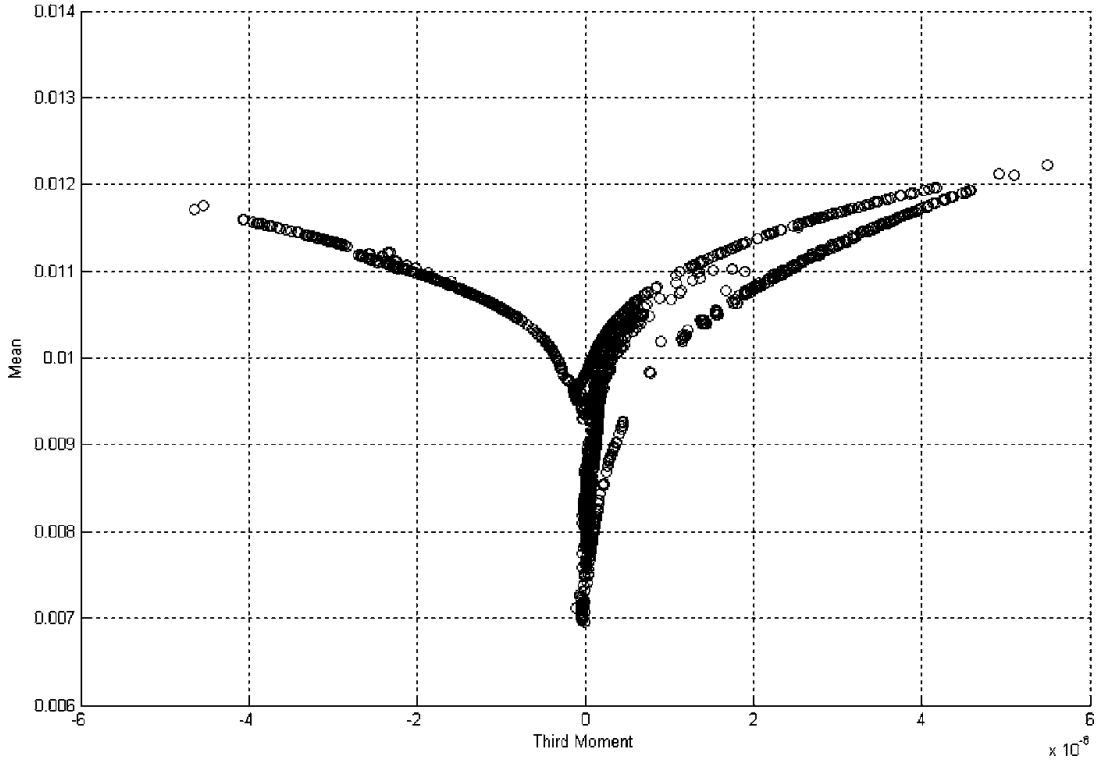


Figure 3.7 Mean–variance–skewness–*kurtosis* constrained efficient portfolios in the mean–third moment plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. Optimal points are obtained after optimisation of 20 hedge funds in 1296 directions.

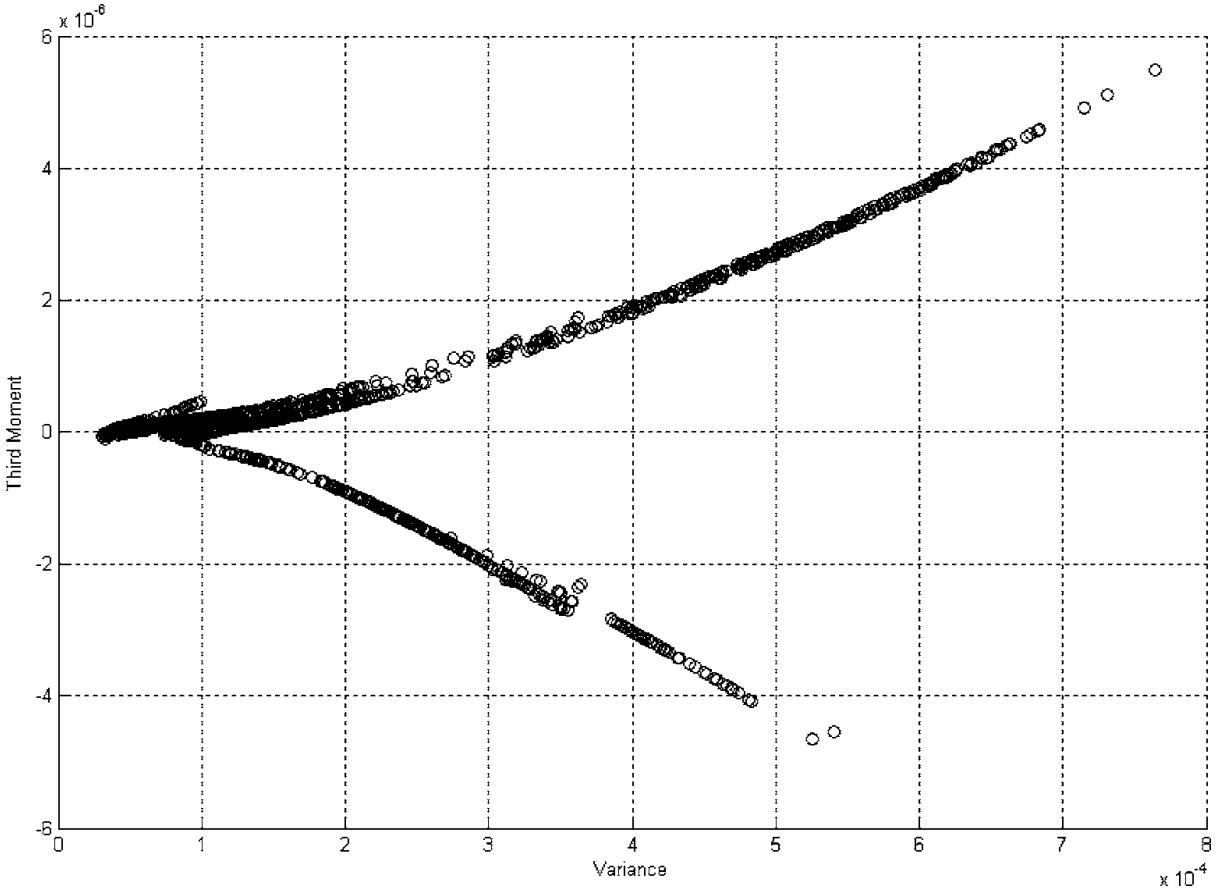


Figure 3.8 Mean–variance–skewness–*kurtosis* constrained efficient portfolios in the variance–third moment plane. *Source*: HFR, monthly net asset values (1995–2005), computations by the authors. Optimal points are obtained after optimisation of 20 hedge funds in 1296 directions.

3.5 CONCLUSION

In this chapter we have introduced a general method for deriving the set of efficient portfolios in the nonconvex mean–variance–skewness–*kurtosis* space, using a shortage optimisation function (see Luenberger, 1995 and Briec *et al.*, 2004 and 2006). The portfolio efficiency is evaluated by looking simultaneously for variance and *kurtosis* contractions and mean and (positive) skewness expansions. This shortage function is linked to an indirect mean–variance–skewness–*kurtosis* utility function. An empirical application on funds of hedge funds provides a three-dimensional representation of the primal nonconvex four-moment efficient portfolio frontier and illustrates the computational tractabilty of the approach.

We approximate the true but unknown mean–variance–skewness–*kurtosis* efficient frontier by a nonparametric portfolio frontier, using an efficiency measure that guarantees global optimality in the four-moment space. In addition, our shortage function can adapt itself to any particular multi-moment asset allocation focusing on return maximisation, skewness maximisation, variance minimisation and *kurtosis* minimisation. Furthermore, dual interpretations are available without imposing any simplifying hypotheses (see Briec *et al.*, 2006). Unfortunately, no global optimal solution can be guaranteed for the indirect mean–variance–skewness–*kurtosis* utility function. These findings indicate that future developments in asset

allocation models should probably focus on developing portfolio optimisation methods using moment-based primal, rather than utility-based dual, optimisation approaches.

A natural extension of our framework is the development of a shortage function excluding any projections on the vertical or horizontal parts of the nonconvex feasible portfolio set and optimising the direction vector in the four moment dimensions.

Another extension of our work is the development of a more robust nonparametric multi-moment efficient frontier. This can be done either by working with robust estimators (see Parkinson, 1980; Brys *et al.*, 2004; Kim and White, 2004 and Ledoit and Wolf, 2003, 2004a and 2004b) of the conventional higher-order moments, by using proper statistical inference tools for the nonparametric efficient frontier (see Simar and Wilson, 2000) or by substituting conventional moment definitions by alternative ones such as L-moments (see Hosking, 1990 and Serfling and Xiao, 2005). Finally, it would be of great interest to use our approach to gauge the performance of hedge funds.

APPENDIX

Let the local solution of the following quartic optimisation programme:

$$\mathbf{w}_p^* = \underset{w_p}{\text{Arg}} \{\text{Max } \delta\}$$

$$\text{s.t.} \begin{cases} E(R_k) + \delta g_E \leq E(R_p) \\ \sigma^2(R_k) - \delta g_\sigma \geq \sigma^2(R_p) \\ s^3(R_k) + \delta g_s \leq s^3(R_p) \\ \kappa^4(R_k) - \delta g'_\kappa \geq \kappa^4(R_p) \\ \mathbf{w}'_p \mathbf{1} = \mathbf{1} \\ \mathbf{w}_p \geq \mathbf{0} \end{cases}$$

where $\mathbf{w}_{p^*}$ represents the $(N \times 1)$ efficient portfolio weight vector that maximises the performance, risk, skewness and *kurtosis* improvement with respect to the ones of the evaluated portfolio in the direction vector $\mathbf{g}$ be (δ^*, w_p^*). Then $(\delta^*, \mathbf{w}_p^*)$ is also a global solution of (3.7).

Proof Let us denote:

$$\begin{aligned} D = \{(\delta, \mathbf{w}_p) \in (IR_+ \times IR^N) : E(R_k) + \delta g_E \leq E(R_p); \\ \sigma^2(R_k) - \delta g_\sigma \geq \sigma^2(R_p); s^3(R_k) + \delta g_s \leq s^3(R_p); \kappa^4(R_k) - \delta g'_\kappa \geq \kappa^4(R_p) \\ \text{with } (\mathbf{w}'_p \mathbf{1}) = \mathbf{1} \text{ and } \mathbf{w}_p \geq \mathbf{0}\} \end{aligned} \tag{3.10}$$

We have:

$$S_{\text{g}}(\mathbf{w}_p) = \text{Max}\{\delta : (\delta, \mathbf{w}_p) \in D\} \tag{3.11}$$

Assume that the couple $(\delta_1, \mathbf{w}_{p1})$ constitutes a local maximum, but is not a global one. In that case, there exists a couple $(\delta_2, \mathbf{w}_{p2}) \in D$ such that:

$$\delta_2 > \delta_1 \tag{3.12}$$

But since $\mathfrak{D}_p$ satisfies the free disposal property, this implies that for all $\delta \in [\delta_1, \delta_2]$, there exists $\mathbf{w}_p \in \mathfrak{F}_p$ such that $(\delta, \mathbf{w}_p) \in D$. Therefore, there does not exist a neighbourhood $V[(\delta_1, \mathbf{w}_{p1}), \varepsilon]$ where $\varepsilon > 0$, such that $\delta_1 \geq \delta$ for all $(\delta, \mathbf{w}_p) \in V[(\delta_1, \mathbf{w}_{p1}), \varepsilon]$. Consequently, if (δ^*, w_p^*) is a local maximum, then it is also a global maximum. ■

ACKNOWLEDGEMENTS

We are grateful to Thierry Chauveau and Thierry Michel for help and encouragement in preparing this work. The content of this chapter engages only its authors and does not necessarily reflect the opinions of their employers.

REFERENCES

Adcock, C. (2003) Asset Pricing and Portfolio Selection Based on the Multivariate Skew-Student Distribution, Working Paper, University of Sheffield, 15 pages.

Agarwal, V. and N. Naik (2004) Risks and Portfolio Decisions Involving Hedge Funds, *Review of Financial Studies* **17**, 63–98.

Andersen, J. and D. Sornette (2001) Have your Cake and Eat it too: Increasing Returns while Lowering Large Risks!, *Journal of Risk Finance* **2**, 70–82.

Athayde, G. and R. Flôres (2002) Portfolio Frontier with Higher Moments: the Undiscovered Country, Discussion Paper EPGE-FGV, 42 pages.

Athayde, G. and R. Flôres (2004) Finding a Maximum Skewness Portfolio: a General Solution to Three-moments Portfolio Choice, *Journal of Economic Dynamics and Control* **28**, 1335–1352.

Berényi, Z. (2001) Performance of Leveraged Asset Funds, Working Paper, University of Munich, 42 pages.

Berényi, Z. (2002) Measuring Hedge Funds' Risks with Moment-based Variance-equivalent Measures, Working Paper, University of Munich, 35 pages.

Briec, W., K. Kerstens and J. Lesourd (2004) Single-Period Markowitz Portfolio Selection, Performance Gauging, and Duality: A Variation on the Luenberger Shortage Function, *Journal of Optimization Theory and Applications* **120**, 1–27.

Briec, W., K. Kerstens and O. Jokung (2006) Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach, *Management Science forthcoming*, 31 pages.

Brys, G., M. Hubert and A. Struyf (2004) A Robust Measure of Skewness, *Journal of Computational and Graphical Statistics* **13**, 996–1017.

Chang, C., T. Pactwa and A. Prakash (2003) Selecting a Portfolio with Skewness: Recent Evidence from US, European, and Latin American Equity Markets, *Journal of Banking and Finance* **27**, 1375–1390.

Chunhachinda, P., K. Dandapani, K. Hamid and S. Prakash (1997) Portfolio Selection and Skewness: Evidence from International Stock Markets, *Journal of Banking and Finance* **21**, 143–167.

Davies, R., H. Kat and S. Lu (2004) Single Strategy Fund of Hedge Funds, Working Paper, ISMA Center, 27 pages.

Davies, R., H. Kat and S. Lu (2005) Fund of Hedge Funds Portfolio Selection: A Multiple-Objective Approach, Working Paper, ISMA Center, 44 pages.

Gamba, A. and F. Rossi (1997) A Three-moment Based Capital Asset Pricing Model, Working Paper, University of Venice, 16 pages.

Gamba, A. and F. Rossi (1998a) A Three-moment Based Portfolio Selection Model, *Rivista di Matematica per le Scienze Economiche e Sociali* **20**, 25–48.

Gamba, A. and F. Rossi (1998b) Mean-Variance-Skewness Analysis in Portfolio Choice and Capital Markets, *Ricerca Operativa* **28**, Special Issue 1998, 5–46.

Goetzmann, W., J. Ingersoll, M. Spiegel and I. Welch (2004) Sharpening Sharpe Ratios, Working Paper, Yale University, 51 pages.

Guidolin, M. and A. Timmermann (2005) Optimal Portfolio Choices under Regime Switching, Skew and Kurtosis Preferences, Working Paper, Federal Reserve Bank of St Louis, 35 pages.

Harvey, C., J. Lietchty, M. Lietchty and P. Müller (2004) Portfolio Selection with Higher Moments, Working Paper, Duke University, 51 pages.
Hosking, J. (1990) L-moments: Analysis and Estimation, *Journal of the Royal Statistical Society Series B* **52**, 105–124.
Ingersoll, J. (1987) *Theory of Financial Decision Making*, Rowman and Littlefield, Ottowa.
Jondeau, E. and M. Rockinger (2003) How Higher Moments Affect the Allocation of Assets, *Finance Letters* **1** (2), 1–5.
Jondeau, E. and M. Rockinger (2006) Optimal Portfolio Allocation Under Higher Moments, *Journal of the European Financial Management Association* **12**, 29–67.
Jurczenko, E. and B. Maillet (2001) The 3-CAPM: Theoretical Foundations and an Asset Pricing Model Comparison in a Unified Framework. In: *Developments in Forecast Combination and Portfolio Choice*, C. Dunis, A. Timmermann and J. Moody (Eds), John Wiley & Sons, Ltd, Chichester, pp. 239–273.
Kim, T. and H. White (2004) On More Robust Estimation of Skewness and Kurtosis, *Finance Research Letters* **1**, 56–73.
Lai, T. (1991) Portfolio with Skewness: A Multiple-objective Approach, *Review of Quantitative Finance and Accounting* **1**, 293–305.
Ledoit, O. and M. Wolf (2003) Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection, *Journal of Empirical Finance* **10**, 603–621.
Ledoit, O. and M. Wolf (2004a) A Well-conditioned Estimator for Large-dimensional Covariance Matrices, *Journal of Multivariate Analysis* **88**, 365–411.
Ledoit, O. and M. Wolf (2004b) Honey, I Shrunk the Sample Covariance Matrix, *Journal of Portfolio Management* **30**, 110–119.
Luenberger, D. (1995) *Microeconomic Theory*, McGraw-Hill.
Markowitz, H. (1952) Portfolio Selection, *Journal of Finance* **7**, 77–91.
Morey, M. and R. Morey (1999) Mutual Fund Performance Appraisals: A Multi-horizon Perspective with Endogenous Benchmarking, *Omega International Journal of Management Science* **27**, 241–258.
Parkinson, M. (1980) The Extreme Value Method for Estimating the Variance of the Rate of Return, *Journal of Business* **53**, 61–65.
Pressacco, F. and P. Stucchi (2000) Linearity Properties of a Three-moments Portfolio Model, *Decisions in Economics and Finance* **23**, 133–150.
Rubinstein, M. (1973) The Fundamental Theorem of Parameter-preference Security Valuation, *Journal of Financial and Quantitative Analysis* **8**, 61–69.
Serfling, R. and P. Xiao (2005) Multivariate L-Moments, Working Paper, University of Texas, 33 pages.
Simaan, Y. (1993) Portfolio Selection and Asset Pricing Three Parameter Framework, *Management Science* **5**, 568–577.
Simar, L. and P. Wilson (2000) A General Methodology for Bootstrapping in Non-Parametric Frontier Models, *Journal of Applied Statistics* **27**, 779–802.
Sun, Q. and X. Yan (2003) Skewness Persistence with Optimal Portfolio Selection, *Journal of Banking and Finance* **27**, 1111–1121.
Wang, S. and Y. Xia (2002) *Portfolio Selection and Asset Pricing*, Springer-Verlag.
Weisman, A. (2002) Informationless Investing and Hedge Fund Performance Measurement Bias, *Journal of Portfolio Management* **28** (4), 80–91.

Annexe 5 :

"A Robust Hybrid DHMM-MLP Modelling of Financial Crises measured by the WhIMS"

En collaboration avec Christophe Boucher et Bertrand Maillet.

Paru dans les actes du colloque du 17^{th} *European Symposium on Artificial Neural Networks Bruges*, 2009.

A Robust Hybrid DHMM-MLP Modelling of Financial Crises measured by the WhIMS

Christophe Boucher[1] Bertrand Maillet[2] and Paul Merlin[3] *

1- A.A.Advisors-QCG (ABN AMRO), Variances and University Paris 1 (CES/CNRS)
christophe.boucher@univ-paris1.fr - France

2- A.A.Advisors-QCG (ABN AMRO), Variances and University Paris 1 (CES/CNRS and EIF)
bmaillet@univ-paris1.fr - France

3- A.A.Advisors-QCG (ABN AMRO), Variances and University Paris 1 (CES/CNRS)
paul.merlin@univ-paris1.fr - France

Abstract. This paper develops a hybrid model combining a Hidden Markov Chain (HMC) and Multilayer Perceptrons (MLP) on the Wavelet-heterogeneous Index of Market Shocks (WhIMS) to identify dynamically regimes in financial turbulences. The WhIMS is an aggregate measure of volatility computed at different frequencies. We estimate the model based on a French market stock index (CAC40 Index) and compare the prediction performance of the HMC-MLP model to classical linear and non-linear models. A state separation of financial disturbances based on the WhIMS and conditional probabilities of the HMC-MLP model is then performed using a Robust SOM.

1 Introduction

Financial markets occasionally experience sudden and large falls in asset prices. Such extreme events constitute a severe threat for the stability of the financial system and international investors not only with the larges losses they cause but also with the deterioration of the risk-return trade-off and the increase of conditional correlation and volatility across the markets (see e.g. Longin and Solnik, 2002).

Following Maillet *et al.* (2004), this paper proposes to identify regimes in financial turbulences, i.e. normal and crisis states. We present a model for the Wavelet-heterogeneous Index of Market Shocks (WhIMS) based on a Hidden Markov Chain (HMC) and Multilayer Perceptrons (MLP). The WhIMS is an easily computable measure that quantifies the intensity of market movements (see Maillet and Michel, 2003, Boucher *et al.*, 2008). In short, its construction is based on an analogy with the so-called Richter scale used for measuring earthquake intensity. The main step consists in applying a Wavelet Packets Sub-band Decomposition constrained Independent Component Analysis (WPSD-cICA) first to decompose the return volatility at different time scales and, secondly,

*We here acknowledge Rachid Bokreta and Caroline Barrault for excellent preliminary research assistance and an active participation in earlier versions. The usual disclaimer applies.

to extract independent factors resulting of the decomposition. Then, we fit volatility extreme values from a Generalized Extreme Value (GEV) based on the L-moment method.

We first present the Autoregressive Hidden Markov Chain Models. We estimate the model based on a French stock market index (CAC40 Index) and compare the prediction performance of the HMC-MLP model to classical linear and non-linear models. A state separation of financial disturbances based on the WhIMS and conditional probabilities of the HMC-MLP model is finally performed using a Robust SOM.

2 AutoRegressive Hidden Markov Chain Models

Over long periods, financial time series exhibit important breaks in their behavior and properties. Abrupt changes may be due to several reasons, such as bankruptcies, burst of bubbles, structural changes in business cycles conditions (*i.e.* disinflation), or wars and related events. One way to capture structural breaks is to use hidden Markov chains. The hidden Markov chain , initially proposed by Baum and Petrie (1966), has been widely applied in various fields, including speech recognition, signal processing, DNA recognition and financial time series.

Let us consider $(Y_t)_{t\in\mathbb{N}}$ the observed time series and let $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ be a homogeneous Markov chain defined by its state space $S = \{s_1, ..., s_N\}$, $N \in \mathbb{N}^\star$ and the $(N \times N)$ transition matrix $\mathbf{A}$ with $a_{ij} = P(\mathbf{X}_{t+1} = s_i \mid \mathbf{X}_t = s_j)$, $(i, j) = [1, ..., N]^2$. If we suppose, with no loss of generality, that the chain state space is the canonical basis of $\mathbb{R}^N$ and we note $\mathbf{v}_{t+1} = \mathbf{X}_{t+1} - E[\mathbf{X}_{t+1} | \mathbf{X}_t]$, then an autoregressive hidden Markov chain model has the following form:

$$\begin{cases} \mathbf{X}_{t+1} = \mathbf{A}\mathbf{X}_t + \mathbf{v}_{t+1} \\ Y_{t+1} = F_{\mathbf{X}_{t+1}}(\mathbf{Y}_{t,t-p+1}) + \sigma_{\mathbf{X}_{t+1}}\, \varepsilon_{t+1} \end{cases} \tag{1}$$

where $\mathbf{Y}_{t,t-p+1}$ defines the vector $(Y_{t-p+1}, ..., Y_t)$, $F_{\mathbf{X}_{t+1}} \in \{F_{s_1}, ..., F_{s_N}\}$ is an autoregressive function of order p, $\sigma_{\mathbf{X}_{t+1}} \in \{\sigma_{s_1}, ..., \sigma_{s_N}\}$ is a real strictly positive number, $(\varepsilon_t)_{t\in\mathbb{N}}$ are independently, identically distributed according to a standardized Gaussian law. In particular, we apply nonlinear autoregressive functions, such as the multilayer perceptrons, to consider the so-called hybrid HMC-MLP models (Hidden Markov Chain - Multilayer Perceptron models).

3 Research of a HMC-MLP Model

We focus on daily values of the WhIMS, computed from the CAC40 French stock market index, from July 9th, 1987 until April 30th, 2008 (5,243 observations). We first investigated whether the series (estimated using the Quasi-Maximum Likelihood method) is "regime switching" and what type of structure should be chosen and, secondly, if we can distinguish different market behaviors by using a hybrid Hidden Markov Chain – Multilayer Perceptron model (HMC-MLP). The linear model and the one hidden-layer perceptron that were considered in a

preliminary study showed that the significant lags were 1, 2, 4, 5 and 6. We fixed this input vector and we let both the number of experts and hidden units vary up to three. In the end, two configurations were selected and further investigated, on the basis of two empirical criteria: the first architecture is the one to have the smallest mean squared error and the second is the one with the "best" transition matrix. In the latter case, the "best" transition matrix is the one empirically generating the most stable segmentation of the series, that is its trace divided by its dimension is the closest to one.

3.1 Prediction Performance

First, we compare the estimated HMC-MLP model with the results of a linear ARMA (see Brockwell *et al.*, 2002) and of a Multilayer Perceptron. After investigating all ARMA(p, q) models, for p and q in a specified range (both lower than 10), an AR(6) model minimizing the BIC criterion was selected. The research of the Multilayer Perceptron is done using the Baum-Welch algorithm (see Baum *et al.*, 1970) for Maximum Likelihood parameter estimation and the Viterbi algorithm (see Viterbi, 1967) to compute the most probable sequence. The "best" model is chosen to minimize the BIC criterion and the "Statistical Stepwise" algorithm is used to eliminate the non-significant connections, once the "dominating" perceptron is found. The final selected model has two hidden units, twelve parameters and the input vector time dimension goes up to 6.

Comparisons of the mean squared error (MSE) suggest there is no special interest in considering a HMC-MLP model for forecasting the exact value of the WhIMS as there is no sensible improvement when compared to the linear model or to the perceptron (results available upon request). Intuitively, this negative result is not surprising when recalling that sudden and violent movements in the market are clearly difficult to predict, and that crises are often due to exogenous and unforecastable events (*e.g.* for the September 2001 terrorist attack). Nevertheless, a state representation of the turbulence is of interest since a qualitative assessment (crisis *versus* non-crisis period) is worthy and decisive for financial applications.

3.2 A Three-state HMC-MLP Model

The research of a three state model having the most significant segmentation of the series leads to select an architecture with the following estimated transition matrix:

$$\hat{A} = \begin{pmatrix} .91 & .06 & .03 \\ .02 & .83 & .15 \\ .02 & .13 & .85 \end{pmatrix}$$

The three experts are three hidden units perceptrons. The associated estimated variances are, respectively, .77, .36 and .24 and the mean squared error is 1.03. The interpretation of a three-state model is however not straightforward than a simple two-state one (crisis *versus* non-crisis), but preliminary results

suggest that a 3-state model allow better discrimination of crisis and non-crisis periods than a 2-state model. We present hereafter a robust Kohonen (2000) classification for describing the behavior of the Wavelet-heterogeneous Index of Market Shocks subject to the three experts. But, since during the learning process the outliers influence the model by moving the neurons toward the outliers, we use a robust version of the SOM (see Maillet *et al.*, 2005).

We consider the WhIMS and the conditional probabilities related to the three perceptrons as input variables. The three probabilities come from the HMC-MLP model and thus correspond to a denoized estimate of the market conditions.

We first conduct a hierarchical classification performed on the map. If we aggregate the information contained in this Kohonen map, by cutting the classification tree and only keep three clusters, we note that the WhIMS values migrate through clusters, from small to high values; the first and the last cluster are homogeneous and correspond, respectively, to periods of calm and crisis respectively. The intermediate cluster - associated to medium WhIMS values and the third expert, but also to mixtures of experts mentioned above - is less homogeneous. The separation between high values of the WhIMS - associated with the first expert - and low values of the Index - linked to the predictions of the second expert - are quite insensitive to the number of clusters. Performing a Kohonen map analysis has here the major advantage that the cut between regimes is less arbitrary, and show that a clear separation can be made based on the WhIMS value and the value of conditional probabilities to be in some states.

Finally, we investigate the behavior of returns in the three identified clusters. Table 1 presents some performance measures of portfolios corresponding to each identified regime. Differences between the three clusters clearly reflect the various disturbance regimes from high return-low volatility (cluster 1) to low return-high volatility (cluster 3), with an intermediate state which is undetermined and corresponds to a wider range of WhIMS. The cumulated differences on the whole sample amount to large discrepancies between conditional performances. Moreover, the state separation based on the WhIMS leads to a better discrimination of market conditions than those based on the other financial disturbance measures.

The three portfolios derived from the WhIMS are shown on Figure 1: the series "State 1" ("State 2", "State 3") is built considering the benchmark returns when the period is classified in the first (second, third) cluster (of the three-category classification) and a zero return when classified elsewhere, either in the second (first, first) clusters or third (third, second). While the first state (first cluster) is clearly associated with low WhIMS and low return, the second cluster is more in line with volatile markets, whilst the third cluster is generally associated with large drops in the market. These results suggest that jumps or regime switching in volatitily could potentially strongly affect portfolio allocation.

4 Conclusion

This paper proposes to model the WhIMS - an index of market disturbances - in a HMC-MLP framework to account for potential regime switching in financial turbulences. A non-linear classification, such a robust Kohonen map analysis, based on the WhIMS and conditional probabilities of the HMC-MLP model allows to identify and characterise stock market conditions. From an asset allocation and risk-management perspective, a promising direction of future research would be to investigate how the identification of market condition regimes based on expert conditional probabilities altogether with WhIMS values affects the investor's portfolio optimisation problem

References

[1] Baum L. E., and T. Petrie, (1966), "Statistical inference for probabilistic functions of finite state Markov chains?", *Annals of Mathematical Statistics 37(6)*, 1554-1563.

[2] Baum L. E., T. Petrie, G. Soules and N. Weiss, (1970), "A Maximization Technique occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *Annals of Mathematical Statistics 41(1)*, 164–171.

[3] Boucher C., B. Maillet and H. Raymond, (2008), "A Wavelet-heterogeneous Index of Market Shocks for assessing the Amplitude of Financial Crises", *CES Working Paper*, forthcoming in 2009, 30 pages.

[4] Brockwell P., R. Davis and P. Rockwell, (2002), *Introduction to Time Series and Forecasting*, Second Edition, Springer-Verlag, New York, 456 pages.

[5] Kohonen T., (2000), *Self-Organizing Maps*, Springer Series in Information Sciences, Vol. 30, Springer, New-York, 501 pages.

[6] Longin F. and B. Solnik. (2002), "Extreme Correlation of International Equity Markets", *Journal of Finance 56(2)*, 649-676.

[7] Maillet B., Merlin P. and P. Rousset (2005), "Robust SOM for Realistic Data Completion", Proceedings of the 5th Workshop On Self-Organizing Maps, 371-378.

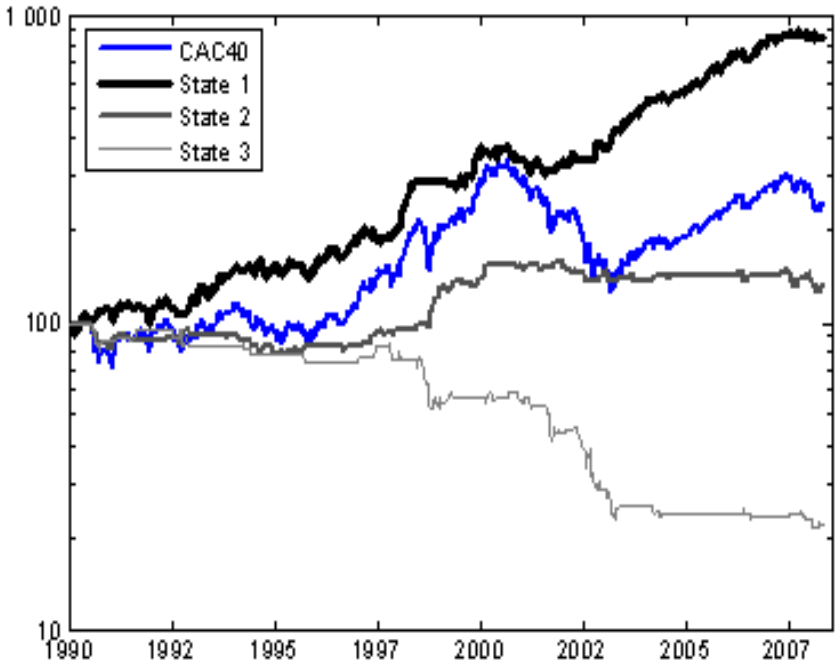
[8] Maillet B. and T. Michel, (2003), "An Index of Market Shocks based on Multiscale Analysis", *Quantitative Finance 3(2)*, 88-97.

[9] Maillet B., Olteanu M. and J. Rynkiewicz (2004), "Non-linear Analysis of Shocks when Financial Markets are Subject to Changes in Regime", Proceedings of ESANN 2004 - European Symposium on Artificial Neural Networks, Bruges, Belgium.

[10] Reinhart C. and C. Rogoff, (2008), "This Time is Different: A Panoramic View of Eight Centuries of Financial Crises", *NBER Working Paper N°13882*, 123 pages.

[11] Viterbi A., (1967),"Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Transactions on Information Theory 13*, 260-269.

Fig. 1: Cumulated Return Series Related to the State of Turbulence of the CAC 40



Source: Bloomberg. Weekly CAC40 Index data between January 1st. 1987 and April 30th. 2008. "State 1" ("State 2". "State 3") corresponds to a series of cumulated returns (base 100 in January 1st. 1990) when the WhIMS and the conditional expert probabilities are classified in the first (second, third) cluster on a three-category string (semi-logarithmic scale). Computations by the authors.

Table 1: Comparison between Market Characterizations based on a HMM-MLP Modelling of the IMS, WhIMS, Implied Standard Deviation or One-year Volatility

	Freq.	**Return**	**Vol.**	**Sharpe Ratio**	**Up**	**Large Down**
All States						
CAC40	100	5.24	20.56	8.46	54.38	10
State 1						
WHIMS	63.68	25.5	15.5	141.88	60.54	4.09
IMS	44.54	20.85	15.58	111.34	59.12	4.14
VIX	35.78	11.36	17.66	44.51	55.76	7.58
Vol.	31.68	15.25	20.7	56.79	57.34	9.56
State 2						
WHIMS	29.84	-15.69	24.43	-78.57	44.93	17.75
IMS	35.03	-2.25	19.84	-28.97	51.85	12.04
VIX	37.08	7.99	15.51	28.96	57.14	6.71
Vol.	36.32	2.28	19.36	-6.3	51.94	7.46
State 3						
WHIMS	6.49	-47.44	35.49	-143.54	38.33	35
IMS	20.43	-11.58	29.26	-51.54	48.68	19.04
VIX	27.14	-5.71	28.62	-32.19	49	17.53
Vol.	32	-0.7	21.71	-19.32	54.39	13.18

Source: Bloomberg. Weekly CAC40 Index data between January 1st, 1987 and April 30th, 2008. Computations by the authors. The IMS corresponds to the Index of Market Shocks (Maillet and Michel, 2003), whilst the WhIMS to the Wavelet-heterogeneous Index of Market Shocks, the VIX the Index of Implied Volatility and the Volatility correspond to the one-year daily annualized volatility of returns on the CAC40. All figures - except Sharpe ratios - are expressed in percentages. The column "State" indicates the regime issued from the classification. Frequency represents the number of period in the state. Mean, volatility and Sharpe ratios represent annualized first and second central conventional moments of the conditional return in the various states. Up (Large down) frequency counts the number of positive (large negative) return in each state conditional samples.

Annexe 6 :

"X-SOM and L-SOM: A Nested Approach for Missing Value Imputation"

En collaboration avec Antti Sorjamaa, Amaury Lendasse et Bertrand Maillet.

Paru dans les actes du colloque du *17th European Symposium on Artificial Neural Networks Bruges*, 2009.

X-SOM and L-SOM: a Nested Approach for Missing Value Imputation

Paul Merlin[1], Antti Sorjamaa[2], Bertrand Maillet[1] and Amaury Lendasse[2]

1- A.A.Advisors-QCG - Variances and University of Paris-1 (CES/CNRS and EIF)
106 bv de l'hôpital F-75647 Paris cedex 13 - France

2- Helsinki University of Technology - ICS
P.O. Box 5400, 02015 HUT - Finland

Abstract. In this paper, a new method for the determination of missing values in temporal databases is presented. This one is based on a robust version of a nonlinear classification algorithm called Self-Organizing Maps and it consists of a combination of two classifications in order to take advantage of spatial as well as temporal dependencies of the dataset. This nested approach leads to a significant improvement of the estimation of the missing values. An application of the determination of missing values for hedge fund return database is presented.

1 Introduction

The presence of missing values in the underlying time series is a recurrent problem when dealing with databases. Because of the absolute need of complete time series for most of the models, a number of methods to handle missing data have been proposed.

Self-Organizing Maps [1] (SOM) aim ideally to group homogeneous individuals through a low-dimensional projection and to highlight the neighborhood structure between the classes. The SOM networks have the ability to be robust, even when some values are missing [2]. SOM-based methods for recovering the missing values have already been proposed, for instance in [3] and [4]. They usually make an intensive use of the spatial correlation and fill the missing values of a time series by the corresponding values of the network neurons after training.

However, one can mention two main drawbacks. First, the dynamics of the time series are not taken fully into account, and secondly, the rebuilding process is discrete. As in [5], we propose, a combination of a transversal (X-SOM) and a longitudinal (L-SOM) classifications allowing us to overcome the above limits and to incorporate spatial, as well as temporal dependencies.

The structure of this paper is as follows. In Section 2, the SOM algorithm and its robust version are presented. The following section is dedicated to present the new algorithm for conditional missing values recovery. In the last section, a financial time series return dataset is used to illustrate the accuracy of the method.

2 Self-Organizing Maps

The SOM algorithm is based on an unsupervised learning principle, where training is entirely data-driven and no information about the input data is required [1]. Here, we use a 2-dimensional network, compound in c units (or code vectors) shaped as a square *lattice*. Each unit of a network has as many weights as the length T of the learning data samples, $\mathbf{x}_n$, for $n = [1, 2, ..., N]$. All units of a network can be collected to a weight matrix $\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_c(t)]$ where $\mathbf{m}_i(t)$ is the T-dimensional weight vector of the unit i at time t and t represents the steps of the learning process. Each unit is connected to its neighboring units through neighborhood function $\lambda(\mathbf{m}_i, \mathbf{m}_j, t)$, which defines the shape and the size of the neighborhood at time t.

First the network nodes are initialized randomly from the data sample space. Then, the iterative learning process begins. For a randomly selected sample $\mathbf{x}_{t+1}$, the Best Matching Unit (BMU), which is the neuron whose weights are closest to the sample is calculated as $\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg\min_{\mathbf{m}_i, i \in I} \{\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|\}$, where $I = [1, 2, ..., c]$ is the set of network node indices, BMU denotes the index of the best matching node and $\|.\|$ is standard Euclidean norm.

If the randomly selected sample includes missing values, the BMU cannot be solved outright. Instead, an adapted SOM algorithm [2] is used. The randomly drawn sample $\mathbf{x_{t+1}}$ having missing value(s) is split into two subsets $\mathbf{x}_{t+1}^T = NM_{\mathbf{x_{t+1}}} \cup M_{\mathbf{x_{t+1}}}$, where $NM_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are not missing and $M_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are missing. We define a norm on the subset $NM_{\mathbf{x_{t+1}}}$ as

$$\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} = \sum_{k \in NM_{\mathbf{x_{t+1}}}} [\mathbf{x}_{t+1,k} - \mathbf{m}_{i,k}(t)]^2, \quad (1)$$

where $\mathbf{x}_{t+1,k}$ denotes the k^{th} value of the chosen data vector and $\mathbf{m}_{i,k}(t)$ is the k^{th} value of the i^{th} code vector. k goes through all the indexes in the subset $NM_{\mathbf{x_{t+1}}}$, where values are not missing.

Then the BMU is calculated with

$$\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg \min_{\mathbf{m}_i, i \in I} \left\{ \|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} \right\}. \quad (2)$$

When the BMU is found the network weights are updated as

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon(t)\lambda\left(\mathbf{m}_{BMU(\mathbf{x}_{t+1})}, \mathbf{m}_i, t\right)[\mathbf{m}_i(t) - \mathbf{x_{t+1}}], \forall i \in I, \quad (3)$$

where $\varepsilon(t)$ is the adaptation gain parameter, which is $]0, 1[$-valued, decreasing gradually with time. The number of neurons taken into account during the weight update depends on the neighborhood function $\lambda(., ., .)$.

After the weight update, the next sample is randomly drawn from the data matrix and the procedure is started again by finding the BMU of the sample. The recursive learning procedure is stopped when the SOM algorithm has converged.

Since our method is able to handle missing values by making an intensive use of the SOM algorithm, issues regarding the SOM convergence have a significant

impact on the missing value reconstruction quality. One way to ensure the convergence is to use the Robust SOM (RSOM) [6]. The idea is to use a bootstrap process to ensure the convergence.

First, an empirical probability for any pair of individuals to be neighbors in the SOM map is estimated with a resampling technique: 40% of observations and individuals are removed, the SOM learning process is performed and finally the removed individuals are projected onto the map, allowing us to get the whole neighborhood structure. The above technique is repeated several times and the empirical estimate of the probability is calculated.

Then the SOM algorithm is, once again, executed several times, but without resampling. From these maps, we select the one whose neighborhood structure is the closest to the empirical probability obtained at the previous step.

The benefits of such a procedure are double. First, the bootstrap process applied during the step one allows the minimization of the effect of possible outliers present in the database. Second, the map chosen in the second step is the one, which maximizes the likelihood of the neighborhood structure.

3 Nested SOM-based Estimation Methodology

SOM-based estimation methods have ever been proposed (for instance, [7] or [3]). These methods typically classified time series and then, using peer-group specificities like mean of individuals or the code vector itself, estimated candidates for the missing values. However, one can mention two main drawbacks. First, the dynamics of the time series are not taken into account, and second, the rebuilding process is discrete, missing values of the time series are filled by the corresponding values of the neurons to which the time series is closes to. Thus, for all series belonging to the same cluster, the estimations are the same.

Following [5], we propose a double classification to overcome the limits. As previously seen in [7] and [3], the first network, identified by its code vector weights $\mathbf{m}^1$ (each unit corresponding to a T-dimensional weight vector), groups individuals, through a longitudinal classification (denoted L-SOM). Then, for each time series $\mathbf{x}_i$ containing missing values, the weights of the associated BMU are substituted for any missing values

$$\mathbf{x}_{i,k} = \mathbf{m}_{BMU(\mathbf{x_i}),k}, \tag{4}$$

for $k \in M_{\mathbf{x}}$.

Simultaneously, we run another SOM classification $\mathbf{m}^2$, on the transversal dataset $\mathbf{x}'$ (each unit corresponds to an n-dimensional weight vector, where n is the number of time series in $\mathbf{x}$). The second cross classification (denoted X-SOM) no more clusterizes observations but realizations. Estimation of missing values operates exactly as in Equation 4.

We have now, two nonlinear estimations for each missing value $\mathbf{x}_{i,k}$ of the dataset. The first one is accurate when considering spatial dependencies, whereas the second integrates temporal correlations more efficiently. We propose to linearly combine these two candidates according to their distances to their respec-

tive BMUs. Let d_1 be the inverse of the distance from the sample $\mathbf{x}_i$ to its associated BMU in $\mathbf{m}^1$, $d_1 = \left(\left\| \mathbf{x_i} - \mathbf{m}^1_{BMU(\mathbf{x_i})} \right\|_{NM_{\mathbf{x_i}}} \right)^{-1}$. We define d_2 equivalently as $d_2 = \left(\left\| \mathbf{x'_k} - \mathbf{m}^2_{BMU(\mathbf{x'_k})} \right\|_{NM_{\mathbf{x'_k}}} \right)^{-1}$.

Then, for each missing value of $\mathbf{x}_{i,k}$, we estimate the missing values contained in the sample through the Nested SOM by

$$\mathbf{x}_{i,k} = d_1 / (d_1 + d_2)\, \mathbf{m^1}_{BMU(\mathbf{x_i}),k} + d_2 / (d_1 + d_2)\, \mathbf{m^2}_{BMU(\mathbf{x'_k}),i}. \tag{5}$$

For the Nested SOM, we still have to select the optimal grid sizes c^1 and c^2. This is done by using validation and the same validation sets for all combinations of the parameters c^1 and c^2. The Nested SOM that gives the smallest validation error is used to perform the final completion of the data.

4 Experimental Results

In the following application, we illustrate our imputation method on a dataset of hedge fund returns[1] composed of 120 funds containing 120 monthly returns from a 10-year period.

Since the hedge fund strategies are well diversified, such assets guarantee us that the time series are not (too much) interdependent. The observed correlations between the assets remain reasonable; the mean, minimum and maximum correlations are respectively .10, $-.62$ and .77. Moreover, since we do not want to favor one of the two classifications (spatial or temporal), we only keep the first 120 funds so that the number of observations remains equal to realizations. Regarding the correlations of the transposed dataset, we find that the mean, minimum and maximum cross correlations are .00, $-.75$ and .74, respectively.

Figure 1 shows 15 among the 120 rescaled fund values[2]. The fund values are low-correlated time series and there are no missing values originally contained in the database.

Before running any experiments, we randomly removed 7.5 percent of the data to a test set. The test set contains 1 080 values. For the validation, the same amount of data is removed from the dataset. Therefore, for the model selection and learning we have a database with a total of 15 percent missing values.

The Monte Carlo Cross-Validation with 20 folds is used to select the optimal parameters for the L-SOM, the X-SOM and the Nested SOM method. The 20 selected validation sets are the same for each method. The validation errors are shown in Figure 2. In the case of the Nested SOM, the errors shown are the minimum errors after the X-SOM with different L-SOM sizes.

The optimal size of the L-SOM grid is found to be 10×10, which is a total of 100 units. We have roughly as many code vectors in the map as observations

[1] provided by *HFR*.

[2] $v'_t = 100 \prod_{i=1}^{t} (1 + r_i)$, with r_i the return of a fund at the time i.

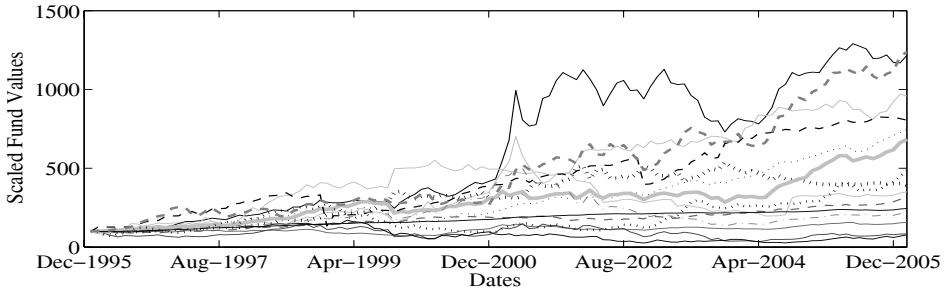


Fig. 1: Rescaled asset values of 15 funds present in the database.

(120). Regarding the cross classification, the X-SOM, we find an optimal size of the grid to be 6×6. It means that we have a nonlinear interpolation between observations and a better approximation of the missing values with more units than data.

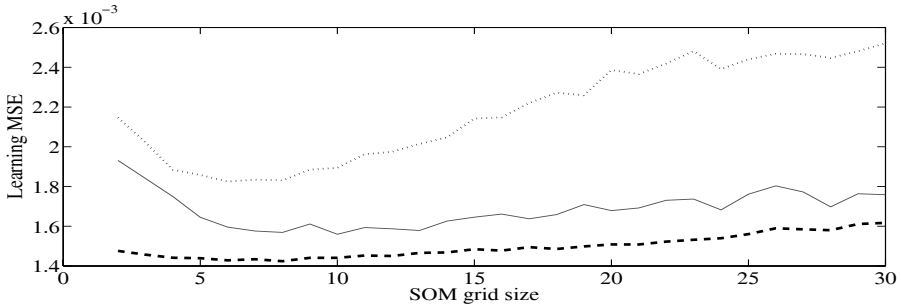


Fig. 2: Validation errors against SOM grid sizes. The L-SOM validation error is the solid line, X-SOM is the dotted line and The Nested SOM is the dashed line.

The smallest error achieved with the Nested SOM method is with the L-SOM grid size 17×17 and with the X-SOM grid size 8×8. The number of neurons is larger for both methods when combined than the L-SOM or X-SOM classification alone. It suggests that the local approximations reduce errors from both L-SOM and X-SOM estimations and enable finer interpolations. From the Figure 2 it is clearly notable that with every SOM size the Nested SOM method gives lower validation error than either L- or X-SOM alone.

Table 1 contains the validation and test errors of all three methods. We can see that the Nested SOM outperforms the L-SOM and the X-SOM reducing the validation error by 19 and 28 percent, respectively, and the test error by 23 and 31 percent.

Table 1: Learning and Test Errors for the L-SOM, the X-SOM and the Nested SOM.

10^{-4}	Learning Error	Test Error
L-SOM	1.6	1.7
X-SOM	1.8	1.9
Nested SOM	1.3	1.3

5 Conclusion

In this paper, we have proposed a new Nested SOM-based method for finding missing values. The L-SOM classification provides efficient missing value estimations that respect spatial dependency structures, whereas the estimations obtained through the X-SOM integrate efficiently the temporal correlations. The combination of these two approaches allows us to overcome the main drawback of the SOM-based imputation methods: the fact that the missing value estimations are discrete. Indeed, considering the distance between series and their associated Best Matching Units make it possible to obtain local continuous approximations of the missing values. As we have shown in the experiments, the combined approach provided estimations that are more accurate than those obtained with each of the methods applied individually.

For further work, the rule driving the local interpolation step may be (easily) upgraded. We also plan to provide a heuristic to find the optimal sizes of the SOMs. The methodology will then be tested on various datasets, and compared to various methodologies to handle missing data. One can also easily think to adopt a conditional approach applying the X-SOM on each L-SOM peer-group to take advantage of the time varying characteristics of the clusters.

References

[1] Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 1995.

[2] Tariq Samad and Steven Harp. Self-organization with partial data. *Network*, 3(2):205–212, 1992.

[3] Marie Cottrell and Patrick Letrémy. Missing values: Processing with the kohonen algorithm. pages 489–496. Applied Stochastic Models and Data Analysis, Brest, France, 17-20 May, 2005.

[4] Antti Sorjamaa, Bertrand Maillet, Paul Merlin, and Amaury Lendasse. Som+eof for finding missing values. pages 115–120. European Symposium on Artificial Neural Networks, Bruges, Belgium, 25-27 April, 2007.

[5] Geoffroy Simon, Amaury Lendasse, Marie Cottrell, Jean-Claude Fort, and Michel Verleysen. Double som for long-term time series prediction. Workshop on Self-Organizing Maps, Kitakyushu, Japan, 11-14 September.

[6] Christiane Guinot, Bertrand Maillet, and Patrick Rousset. Understanding and reducing variability of som neighbourhood structure. *Neural Networks*, 19(6):838–846, 2006.

[7] Françoise Fessant and Sophie Midenet. Self-organising map for data imputation and correction in surveys. *Neural Computing & Applications*, 10(4):300–310, 2002.

Annexe 7 :

"A Non-linear Approach for Completing Missing Values in Temporal Databases"

En collaboration avec Antti Sorjamaa, Amaury Lendasse et Bertrand Maillet.

A paraître dans *European Journal of Economic and Social Systems*, 2009.

A Non-linear Approach for Completing Missing Values in Temporal Databases

Antti Sorjamaa[*] — Paul Merlin[**] — Bertrand Maillet[***] — Amaury Lendasse[*]

** Helsinki University of Technology*
Laboratory of Computer and Information Science
P.O. Box 5400, 02015 HUT - Finland

Antti.Sorjamaa@hut.fi, Lendasse@hut.fi

*** A.A.Advisors (ABN AMRO), Variances and University of Paris-1 (CES/CNRS)*
106 bv de l'hôpital F-75647 Paris cedex 13 - France

paul.merlin@univ-paris1.fr

**** A.A.Advisors-QCG (ABN AMRO), Variances and University of Paris-1 (CES/CNRS and EIF), 106 bv de l'hôpital F-75647 Paris cedex 13 - France*

bmaillet@univ-paris1.fr

ABSTRACT. *The presence of missing data in the underlying time-series is a recurrent problem for market models. Such models impose to deal with cylindrical and complete samples. This paper presents a new procedure for the missing values recovery. The proposed method is based on two projection algorithms: a non-linear one (Self-Organizing Maps) and a linear one (Empirical Orthogonal Functions). The presented global methodology combines the advantages of both methods to get accurate approximations for the missing values. The methods are applied to three financial datasets.*

RÉSUMÉ. *L'absence de certaines valeurs dans les séries temporelles est un problème récurrent lors de l'utilisation de modèles financiers. En effet, de tels modèles nécessitent des bases cylindrées et complètes. Ce papier présente une nouvelle approche pour le recouvrement des valeurs manquantes. Cette méthode utilise deux techniques de projection : une non-linéaire (Cartes de Kohonen) et une linéaire (Fonction Orthogonale Empirique). La méthodologie globale présentée combine les avantages des deux méthodes pour obtenir des candidats aux valeurs manquantes. La méthode est appliquée à trois bases de données financières.*

KEYWORDS: *Missing values, Self-Organizing Maps, Empirical Orthogonal Functions.*

MOTS-CLÉS : *Valeurs manquantes, Cartes de Kohonen, Fonctions orthogonales empiriques.*

1. Introduction

Academics as well as practitioners often face the problem of missing data in financial time-series. Non-quotation date, too recent inception date, intention not to report a bad performance or mistake of data provider are some of the reasons why missing values occur recurrently in financial databases. Moreover, in order to achieve good performance, most financial models need complete and cylindrical samples. Thus, most of the time, imputation methods have to be applied before running the model.

A number of methods, both commercial and academic, have been developed to solve the problem and fill the missing values. They can merely be classified into two distinct categories: deterministic and stochastic methods.

In the later group, the Self-Organizing Maps [KOH 95] (SOM) aim to ideally group homogeneous individuals, highlighting a neighborhood structure between classes in a chosen lattice. The SOM algorithm is based on unsupervised learning principle where the training is entirely stochastic and data-driven. No information about the input data is required. Recent approaches propose to take advantage of the homogeneity of the underlying classes for data completion purposes [WAN 03]. Furthermore, the SOM algorithm allows projection of high-dimensional data to a low-dimensional grid. Through this projection and focusing on its property of topology preservation, SOM allows us for a non-linear interpolation for completing the missing values.

Empirical Orthogonal Functions (EOF) [PRE 88] are deterministic models, enabling a linear projection to high-dimensional space. They have also been used to develop models for finding missing data [BOY 94]. Moreover, EOF models provide continuous interpolation of missing values, but have the disadvantage to be sensitive to the initialization.

This paper describes a new method, that combines the advantages of both the SOM and the EOF. The nonlinearity property of the SOM is used as a denoising tool and then continuity property of the EOF method serves for efficiently recoverying the missing data.

The SOM is presented in Section 2, the EOF in Section 3 and the global methodology SOM+EOF in Section 4. Section 5 presents the experimental results using three financial datasets.

2. Self-Organizing Map

The SOM algorithm is based on an unsupervised learning principle, where training is entirely data-driven and no information about the input data is required [KOH 95]. Here we use a 2-dimensional network, compound in c units (or code vectors) shaped as a square *lattice*. Each unit of a network has as many weights as the length T of the learning data samples, $\mathbf{x}_n$, with $n = [1, 2, ..., N]$. All units of a network can be collected to a weight matrix $\mathbf{m}(t) = [\mathbf{m}_1(t), \mathbf{m}_2(t), ..., \mathbf{m}_c(t)]$ where $\mathbf{m}_i(t)$ is the

T-dimensional weight vector of the unit i at time t and t represents the steps of the learning process. Each unit is connected to its neighboring units through neighborhood function $\lambda(\mathbf{m}_i, \mathbf{m}_j, t)$, which defines the shape and the size of the neighborhood at time t. Neighborhood can be constant through the entire learning process or it can change during the course of learning.

Learning starts by randomly initializing the network node weights. Then, for randomly selected sample $\mathbf{x}_{t+1}$, we calculate a Best Matching Unit (BMU), which is the neuron whose weights are closest to the sample. The BMU calculation is defined as

$$\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg \min_{\mathbf{m}_i, i \in I} \{\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|\}, \quad [1]$$

where $I = [1, 2, ..., c]$ is the set of network node indices, $BMU(.)$ denotes the index of the best matching node and $\|.\|$ is standard Euclidean norm.

If the randomly selected sample includes missing values, the BMU cannot be solved outright. Instead, an adapted SOM algorithm [SAM 92] is used. The randomly drawn sample $\mathbf{x_{t+1}}$ having missing value(s) is split into two subsets $\mathbf{x}_{t+1}^T = NM_{\mathbf{x_{t+1}}} \cup M_{\mathbf{x_{t+1}}}$, where $NM_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are not missing and $M_{\mathbf{x_{t+1}}}$ is the subset where the values of $\mathbf{x_{t+1}}$ are missing. We define a norm on the subset $NM_{\mathbf{x_{t+1}}}$ as

$$\|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} = \sum_{k \in NM_{\mathbf{x_{t+1}}}} (\mathbf{x}_{t+1,k} - \mathbf{m}_{i,k}(t))^2, \quad [2]$$

where $\mathbf{x}_{t+1,k}$ for $k = [1, ..., T]$ denotes the k^{th} value of the chosen vector and $\mathbf{m}_{i,k}(t)$ for $k = [1, ..., T]$ and for $i = [1, ..., c]$ is the k^{th} value of the i^{th} code vector.

Then the BMU is calculated with

$$\mathbf{m}_{BMU(\mathbf{x}_{t+1})} = \arg \min_{\mathbf{m}_i, i \in I} \left\{ \|\mathbf{x_{t+1}} - \mathbf{m}_i(t)\|_{NM_{\mathbf{x_{t+1}}}} \right\}. \quad [3]$$

When the BMU is found the network weights are updated as for each $i \in I$

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) - \varepsilon(t)\lambda\left(\mathbf{m}_{BMU(\mathbf{x}_{t+1})}, \mathbf{m}_i, t\right)[\mathbf{m}_i(t) - \mathbf{x_{t+1}}], \quad [4]$$

where $\varepsilon(t)$ is the adaptation gain parameter, which is $]0, 1[$-valued, decreasing gradually with time. The number of neurons taken into account during the weight update depends on the neighborhood function $\lambda(\mathbf{m}_i, \mathbf{m}_j, t)$. The number of neurons, which need the weight update, usually decreases with time.

After the weight update the next sample is randomly drawn from the data matrix and the procedure started again by finding the BMU of the sample. The recursive learning procedure is stopped when the SOM algorithm has converged.

Once the SOM algorithm has converged, we obtain some clusters containing our data. Fessant and Midenet [FES 02] proposed to fill the missing values of the dataset by the coordinates of the code vectors of each BMU as natural first candidates for missing value completion

$$\pi_{(M_{\mathbf{x}})}\left(\mathbf{x}\right) = \pi_{(M_{\mathbf{x}})}\left[\mathbf{m}_{BMU(\mathbf{x})}\right], \qquad [5]$$

where $\pi_{(M_{\mathbf{x}})}\left(.\right)$ replaces the missing values $M_{\mathbf{x}}$ of sample $\mathbf{x}$ with the corresponding values of the BMU of the sample. The replacement is done for every data sample and then the SOM has finished filling the missing values in the data.

The procedure is summarized in Table 1. There is a toolbox available for performing the SOM algorithm in [URL 01].

Table 1. *Summary of the SOM algorithm for filling the missing values.*

1	SOM node weights are randomly initialized
2	SOM learning process begins
3	Input $\mathbf{x}$ is drawn from the learning data set $\mathbf{X}$
	3.1 If $\mathbf{x}$ does not contain missing values, BMU is found according to Equation 1
	3.2 If $\mathbf{x}$ contains missing values, BMU is found according to Equation 3
4-	Once the learning process is done, for each observation containing missing values, the weights of the BMU of the observation are substituted to missing values

3. Empirical Orthogonal Functions

This section presents Empirical Orthogonal Functions (EOF) [PRE 88]. In this paper, EOF are used as a denoising tool and for finding the missing values at the same time [BOY 94].

The EOF are calculated using standard and well-known Singular Value Decomposition (SVD)

$$\mathbf{X} = \mathbf{UDV}^* = \sum_{k=1}^{K} \rho_k \mathbf{u}_k \mathbf{v}_k, \quad [6]$$

where $\mathbf{X}$ is 2-dimensional data matrix, $\mathbf{U}$ and $\mathbf{V}$ are collections of singular vectors $\mathbf{u}$ and $\mathbf{v}$ in each dimension respectively, $\mathbf{D}$ is a diagonal matrix with the singular values ρ in its diagonal and K is the smaller dimension of $\mathbf{X}$ (or the number of nonzero singular values if $\mathbf{X}$ is not full rank). The singular values and the respective vectors are sorted in decreasing order.

When EOF are used to denoise the data, not all singular values and vectors are used to reconstruct the data matrix. Instead, it is assumed that the vectors corresponding to larger singular values contain more data with respect to the noise than the ones corresponding to smaller values [PRE 88]. Therefore, it is logical to select q largest singular values and the corresponding vectors and reconstruct the denoised data matrix using only them.

In the case where $q < K$, the reconstructed data matrix is obviously not the same than the original one. The larger the q, the more original data is preserved. On the opposite, the smaller the q, the more different is the new data (but the less accurate it is). The optimal q is selected using validation methods, for example [LEN 03].

EOF (or SVD) cannot be directly used with databases including missing values. The missing values must be replaced by some initial values in order to use the EOF. This replacement can be for example the mean value of the whole data matrix $\mathbf{X}$ or the mean in one direction, row wise or column wise. The latter approach is more logical when the data matrix has some temporal or spatial structure in its columns or rows.

After the initial value replacement the EOF process begins by performing the SVD and the selected q singular values and vectors are used to build the reconstruction. In order not to lose **any** information, only the missing values of $\mathbf{X}$ are replaced with the values from the reconstruction. After the replacement, the new data matrix is again broken down to singular values and vectors with the SVD and reconstructed again. The procedure is repeated until convergence criterion is fulfilled.

The procedure is summarized in Table 2.

4. Global Methodology

The two methodologies presented in the previous two sections are combined and the global methodology is presented. The SOM algorithm for missing values is first ran through performing a nonlinear projection for finding the missing values. Then, the result of the SOM estimation is used as initialization for the EOF method. The global methodology is summarized in Figure 1.

Table 2. *Summary of the EOF method for finding missing values.*

1	Initial values are substituted into missing values of the original data matrix $\mathbf{X}$
2	For each q from 1 to K
	2.1 SVD algorithm calculates q singular values and eigenvectors
	2.2 A number of values and vectors are used to make the reconstruction
	2.3 The missing values from the original data are filled with the values from the reconstruction
3	The q with the smallest validation error is selected and used to reconstruct the final filling of the missing values in $\mathbf{X}$

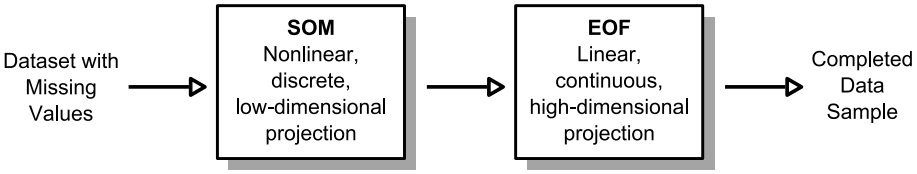


Figure 1. *Summary of the (SOM+EOF) Global methodology.*

For the SOM we must select the optimal grid size c and for the EOF the optimal number of singular values and vectors q to be used. This is done using the same validation set for all combinations of the parameters c and q. Finally, the combination of SOM and EOF that gives the smallest validation error is used to perform the final filling of the data.

Even the SOM as well as the EOF are able to fill the missing values alone, the experimental results demonstrate that together the accuracy is better. The fact that these two algorithms suit well together is not surprising. Two approaches can be considered to understand the complementarity of the algorithms.

Firstly, the SOM algorithm allows nonlinear projection. In this sense, even for dataset with complex and nonlinear structure, the SOM code vectors will succeed to capture the nonlinear characteristics of the inputs. However, the projection is done on a low-dimensional grid (in our case two-dimensional) with the possibility of losing the intrinsic information of the data.

The EOF method is based on a linear transformation using the Singular Value Decomposition. Because of the linearity of the EOF approach, it will not reflect the non-linear structures of the dataset, but the projection space can be as high as the dimension of the input data and remain continuous.

A toolbox for performing the SOM+EOF is available in [URL 02].

5. Experimental Results

To illustrate the efficiency of the presented methodology, we run several experiments on three financial return databases. The first one recovers the missing values when they are missing at random, the second experiment has missing values only at the beginning of several time-series and the third one is a publicly available financial dataset used with random missing values.

In comparison, we also experiment with a widely used methodology called Expectation Conditional Maximization, which is briefly presented in the following.

5.1. *Expectation Maximization Methods*

As a benchmark to estimate how well our combination methodology performs, we choose to compare our results with those obtained by the Expectation Maximization (EM) algorithm.

The EM algorithm presented by Dempster, Laird and Rubin in [DEM 77], is a technique to find maximum likelihood estimates in a missing data situation. Since the estimates of the mean and the covariance matrix of an incomplete dataset depend on the unknown missing values, and, conversely, estimates of the missing values depend on the unknown statistics of the data. This estimation problem is non-linear and has to be done iteratively.

The EM algorithm consists of two steps:

1) E-step calculates the expectation of the complete data sufficient statistics given the observed data and current parameter estimates.

2) M-step updates the parameter estimates through the maximum likelihood approach based on the current values of the complete sufficient statistics.

The algorithm proceeds in an iterative manner until the difference between the last two consecutive parameter estimates converges to a specified *criterion*. The final E-step computes the expectation of each missing value given the final parameter estimates and the observed data. This result will be used as the imputation value.

For each iteration (t), the E-step consists of

$$Q\left(\theta \left| \theta^{(t)} \right.\right) = E\left[L\left(\theta \left| Y \right.\right) \left| Y_{obs}, \theta^{(t)} \right.\right], \qquad [7]$$

where

$$\begin{cases} L\left(.\,|Y\right) \text{ denotes the likelihood function conditionally to the sample,} \\ \theta \text{ the vector of parameter to be estimated,} \\ Y_{obs} \text{ the non-missing values,} \\ Y \text{ the sample,} \\ \theta^{(t)} \text{ the last vector of estimated parameter.} \end{cases}$$

Thus, the $(t+1)^{th}$ M-step finds $\theta^{(t+1)}$ that maximizes $Q\left(\theta \left|\theta^{(t)}\right.\right)$ such that

$$Q\left(\theta^{(t+1)} \left|\theta^{(t)}\right.\right) = \max_{\theta} Q\left(\theta \left|\theta^{(t)}\right.\right). \quad [8]$$

The main drawback of the EM algorithm is when the M-step is not in a closed form. In this case, the M-step could be difficult to perform.

Meng and Rubin [MEN 93] proposed an alternative algorithm called the Expectation Conditional Maximization (ECM) to solve this problem. The M-step is decomposed in multiple conditional maximization. Consider $\theta = [\theta_1, \theta_2, ..., \theta_k]$ a k-dimensional vector of parameters. Then the CM-step consists of k successive maximizations, for $i = 1, ..., k$, with previous notations

$$Q\left(\theta^{(t+1)} \left|\theta^{(t)}\right.\right) = \max_{\theta_i} Q\left(\theta \left|\theta^{(t)}\right.\right). \quad [9]$$

Otherwise, the ECM algorithm performs in the same way than the EM algorithm presented before.

5.2. *North-American Fund Returns*

For the first experiment, we use a dataset of North American fund returns [1] composed with 679 funds on a 4-year period of 219 weekly values, which give a total of 148,701 values. Then, in the definition of the dataset $\mathbf{X}$, the dimensions of the matrix $\mathbf{X}$ is $T \times N$ which is equal to 219×679.

The fund return correspond to the yield of asset values between two consecutive dates as

$$r_t = \frac{v_{t+1}}{v_t} - 1, \quad [10]$$

1. Data provided by Lipper, A Reuters Company.

where v_t is the value of the considered asset at time t.

There are no missing values contained in the original database. Figure 2 shows 10 rescaled fund values $\left(v_t^{'} = 100 \prod_{i=1}^{t} (1 + r_t)\right)$. The fund values are correlated time-series including first order trends.

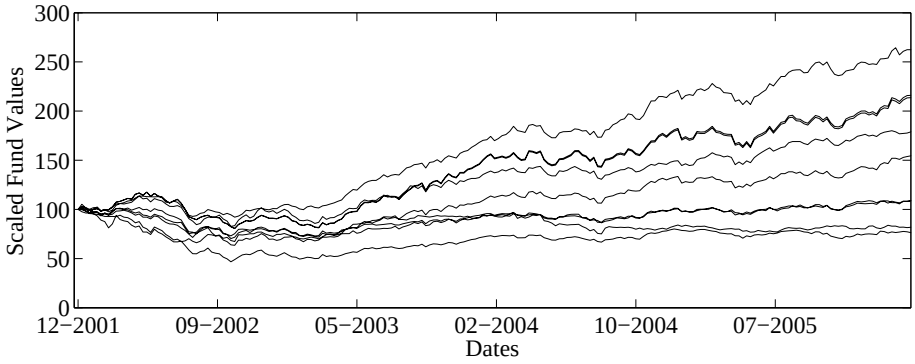


Figure 2. *Rescaled asset values of 10 funds present in the database.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

Before running any experiments, we randomly remove for testing purposes 7.5% of the data, which corresponds to 11,152 missing values. For each validation set, the same amount of data is removed from the dataset. Therefore, for the model selection and learning we have a database with a total of 15% of missing values.

We use Monte Carlo Cross-Validation method with 10 folds to select the optimal parameters for the SOM, the EOF and the SOM+EOF. The 10 selected validation sets are the same for each method and the validation results are presented in the following.

5.2.1. *SOM*

Focusing on the topology preservation property of the SOM algorithm, we project our data on a large sized map. For each grid size, we compute the Root Mean Square Errors (RMSE) of the reconstruction on all validation sets. Then the grid size giving the smallest validation error is selected and the corresponding grid size is used to make the final filling. The validation errors are shown in Figure 3.

The optimal size of the SOM grid is found to be 26×26, which is a total of 676 units, see Figure 3. Therefore, we have more code vectors in the SOM than observations (629). It means that we have a nonlinear interpolation between the observations and better approximation of the missing values.

Once the optimal grid size is found, we apply the SOM algorithm and fill in all the missing values. Now we have only 7.5% of the data missing due to the removed test set. The test and validation errors are summarized in the end of the section, in Table 3.

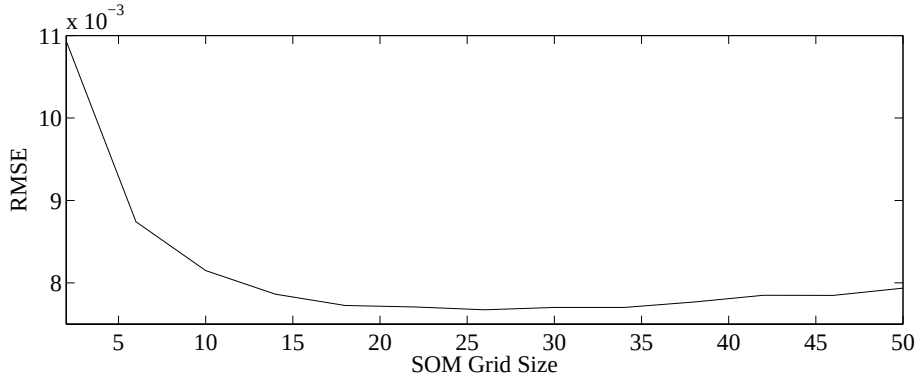


Figure 3. *Validation errors with respect to squared number of grid size using the SOM method.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

5.2.2. *EOF*

The validation errors with respect to q for the EOF method are shown in Figures 4 and 5. In this case, when the EOF is used alone, the missing values are initialized using the column mean of the dataset calculated with only known values of each column.

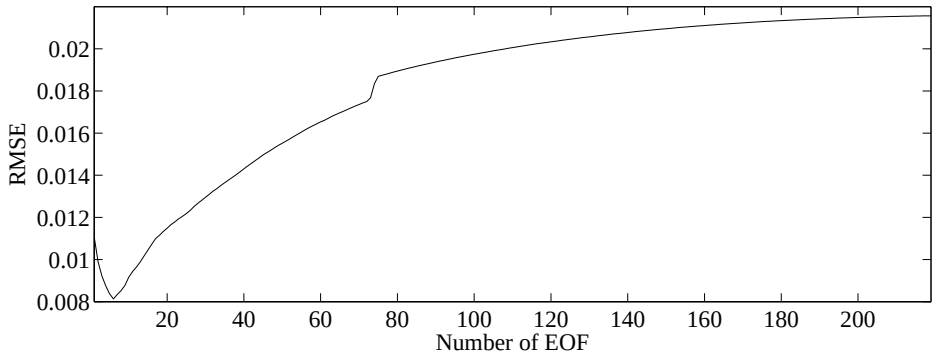


Figure 4. *Validation errors with respect to the number of EOF with the plain column mean initialization.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

From the Figure 5, the smallest error is achieved with q equal to 6. This number of EOF is relatively small compared to the maximum of 219 EOF. It suggests quite strong noise influence in the data and that there is only a small number of efficient EOF needed to represent the denoised data.

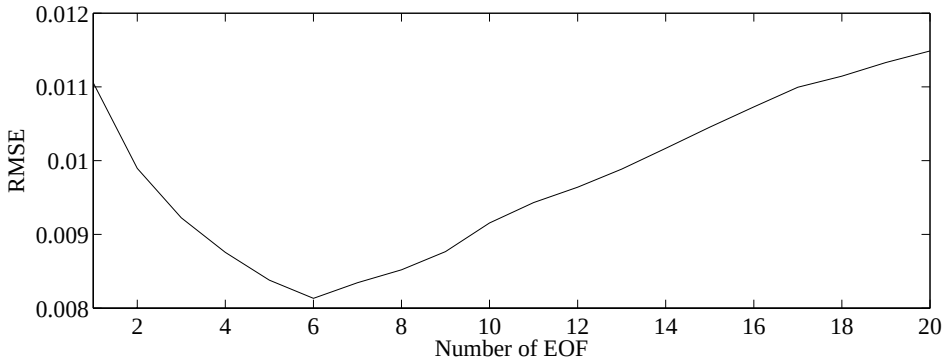


Figure 5. *EOF validation errors zoomed.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

5.2.3. *SOM+EOF*

In our experiments, we have seen that it is not enough to select the SOM grid size and the number of EOF separately. Instead, both parameters must be optimized together, simultaneously. Even though this increases the computational load, it gives more accurate results.

In Figures 6 and 7 the validation RMSEs are presented. The first figure shows the minimum EOF errors with respect to the SOM grid size and the latter figure the EOF errors with the selected SOM grid size.

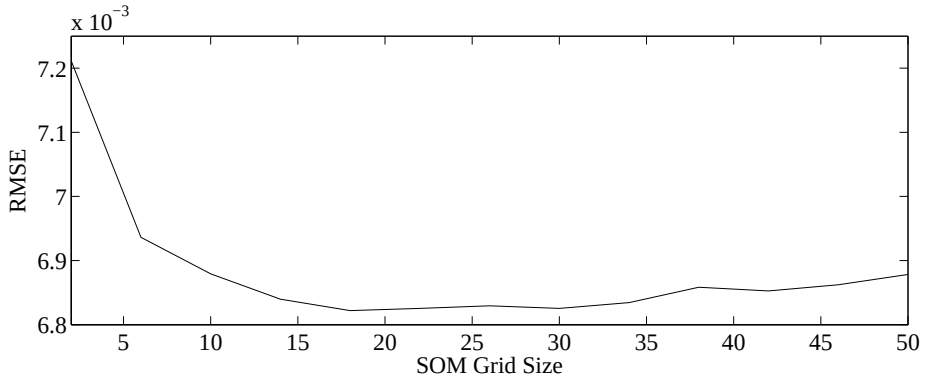


Figure 6. *Validation errors with respect to the SOM grid size using the SOM+EOF.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

From the Figures 6 and 7 the smallest errors are achieved with the SOM grid size equal to 18×18 and the number of EOF q equal to 40.

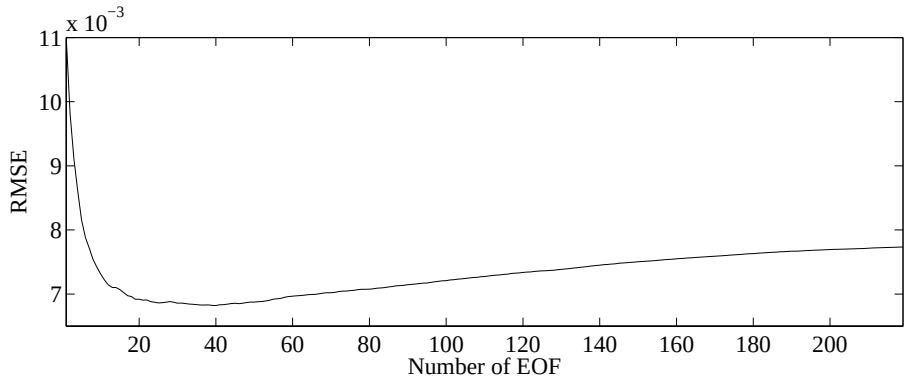


Figure 7. *Validation errors with respect to the number of EOF with the SOM grid size 18×18.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

The number of selected EOF is larger with SOM initialization than with the column mean initialization. It suggests there are more efficient EOF to use in the approximation of the missing values than with the plain column mean initialization and that the SOM has already denoised the data.

The SOM size is decreased when compared to the SOM method alone. It suggests that the nonlinear interpolation is not as crucial than using the SOM alone, but instead the denoising property is enhanced by limiting the number of SOM nodes.

It is also evident that the individual optimization of the parameters does not guarantee appropriate performance, which can be seen from totally different selections of parameters when using the SOM+EOF than the methods individually.

Table 3 summarizes the errors of the SOM, the EOF and the the SOM+EOF methods.

Table 3. *Validation and test RMS errors for all the methods.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

10^{-3}	Validation Error	Test Error
ECM	13.8	13.6
SOM	7.67	7.33
EOF	8.13	7.83
SOM+EOF	6.82	6.59

From the Table 3, we can see that the SOM+EOF outperforms the EOF reducing the validation and test errors by 16% and the SOM errors more than 10%.

5.2.4. *More Missing Values*

In order to test the robustness of the (SOM+EOF) method, we experiment the effect of increasing the percentage of missing values in the database.

Before selecting the test or the validation sets, we randomly remove 30% of the data. Then the same procedure as before is performed by first removing 7.5% of the remaining data for the test set and then for each validation set an extra 7.5% is deleted.

Finally, the total amount of missing data in the learning phase is around 42%, which makes the missing value problem considerably harder than in the previous experiments.

The validation RMS errors for the SOM method are shown in Figure 8.

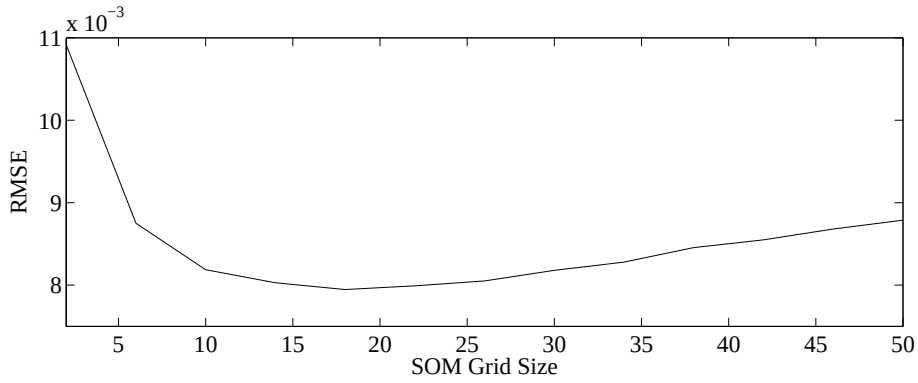


Figure 8. *Validation errors with respect to square number of grid size using the SOM method.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

From Figure 8, the SOM grid size with the smallest RMS error is 18×18, which is smaller than previously using the SOM method. It means that when the percentage of missing values increases, the need for the SOM nodes decrease as there is less data to use in the interpolation of the missing values.

The validation errors for the SOM+EOF method are presented in Figures 9 and 10.

From Figure 9, the optimal SOM size is selected to 18×18, which is the same size than using the SOM alone. It means, that the SOM method alone is definitely not accurate enough to perform the filling of missing values alone. Therefore, it is not possible to enhance the noise removal power over interpolation performance in this case.

From Figure 10, the optimal number of EOF is found to be 15, which is less than in the case with less missing values. The smaller number of EOF is explained by the fact that the increased number of missing values creates more uncertainty and, therefore,

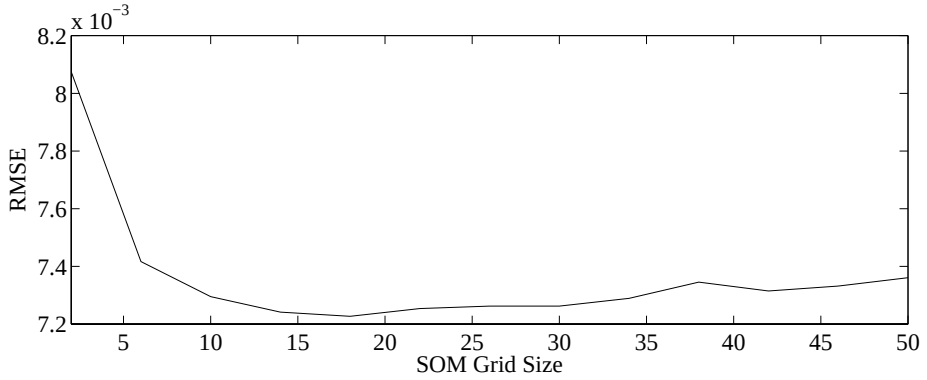


Figure 9. *Validation errors with respect to the SOM grid size using the SOM+EOF.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

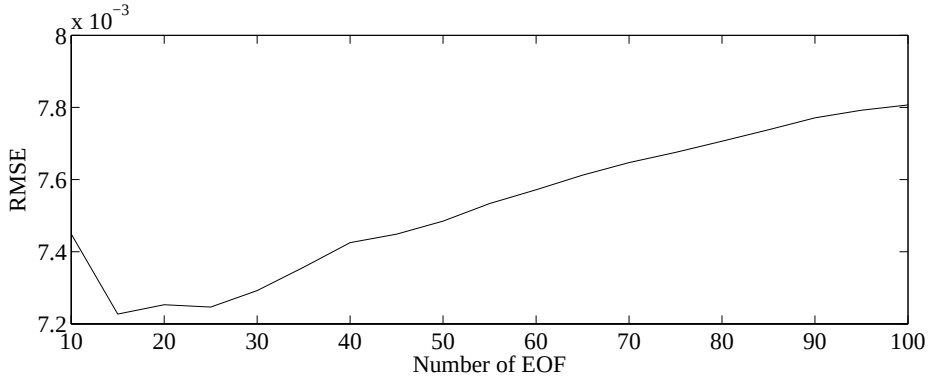


Figure 10. *Validation errors with respect to the number of EOF with the SOM grid size 18×18.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

the smaller singular values and the related vectors become more and more unusable in the reconstruction process.

The validation and test errors are summarized in Table 4.

From Table 4, it can be seen that the SOM+EOF method has decreased the validation and test errors both by 9% compared to the SOM method. The improvement is slightly worse than in the case of less missing values. Still, there is notable performance upgrade when using the SOM+EOF method.

Table 4. *Validation and test RMS errors for all the methods.* *Source: Lipper; North American Fund Weekly Return from 28/12/2001 to 03/03/2006. Computation from the Authors.*

10^{-3}	Validation Error	Test Error
ECM	9.50	13.8
SOM	7.94	7.73
EOF	9.07	9.17
SOM+EOF	7.22	7.01

Comparing the error values above with the values in Table 3, we can see that all errors are increased roughly the same amount, except the ECM method. The validation error is significantly better with more missing values than with less. However, the test error is higher in comparison with the previous case as well as with the SOM, EOF and SOM+EOF methodologies.

Based on the findings, it can be concluded that the SOM and EOF based filling methods are robust and can handle efficiently even large amount of missing values contained in the database.

5.3. *European Fund Returns*

For the next experiment, we focus on a different example of missing values. Rebuilding past performance of funds is a recurrent problem for financial professionals (too short funds history). Thus, we choose to rebuild the beginnings of several time-series. We use a dataset of European Fund Weekly Returns [2] from 07/11/2003 to 27/10/2006 composed of 300 funds with 175 weekly values, which give a total of 52 500 values.

We randomly remove for testing purposes 10% of the data at the beginning of several time-series. The beginning is defined as the first third of the length of the series. We constraint the random deletion process to leave at least one fourth of the time-series without any missing values. For validation, 10

We apply the same validation procedures as in the previous experiments to select the optimal SOM grid size and the number of EOF.

The optimal size of the SOM grid is found to be in mean (12×12) that is 144. Once the optimal grid size is found, we apply the SOM algorithm and fill in all the missing values. When the EOF is performed, initial missing values are substituted as the column means of the original matrix. At last, SOM estimations are then used as initialization for the EOF algorithm.

2. Data provided by Standard and Poors

The validation errors with respect to q for the EOF alone and the EOF performed with the SOM initialization are shown in Figure 11.

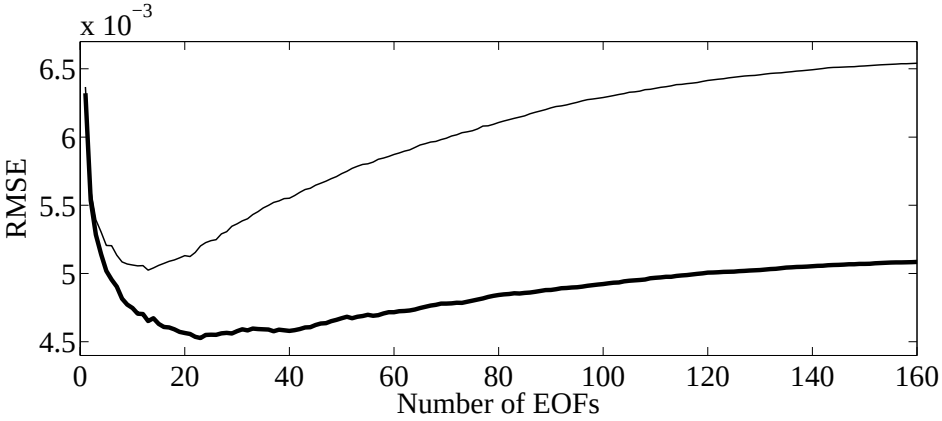


Figure 11. *Validation errors with respect to the number of EOF with the plain column mean initialization, the fine gray line, and with the SOM initialization, the bold black line.* *Source: SnP; European Fund Weekly Return (07/11/2003 to 27/10/2006). Computation from the Authors.*

From the Figure 11, we note that the smallest error is achieved with q equal to 13 using the EOF with plain column mean initialization and 23 when the EOF is initialized using the SOM. Table 5 summarizes the mean errors of the three methods.

Table 5. *Validation and test RMS errors for all the methods.* *Source: SnP; European Fund Weekly Return from 07/11/2003 to 27/10/2006. Computation from the Authors.*

10^{-3}	Validation Error	Test Error
ECM	7.98	8.25
SOM	5.09	5.06
EOF	5.02	4.85
SOM + EOF	4.53	3.83

From the Table 5 we can see that the SOM+EOF outperforms the EOF and the SOM reducing the test error by 31% compared to the SOM and 26% compared to the EOF.

5.4. *Hedge Fund Return Dataset*

The third example is performed with a ten-year monthly return dataset. From a large internal database, we select 100 hedge funds with various strategies (we made it publicly available by removing dates and fund names; it can be found from [URL 02]).

The dataset contains 120 series, 121 values each, and it has no missing values inherently in it. The procedure is the same than before and the results are shown in Figures from 12 to 15 and the validation and test errors are summarized in Table 6.

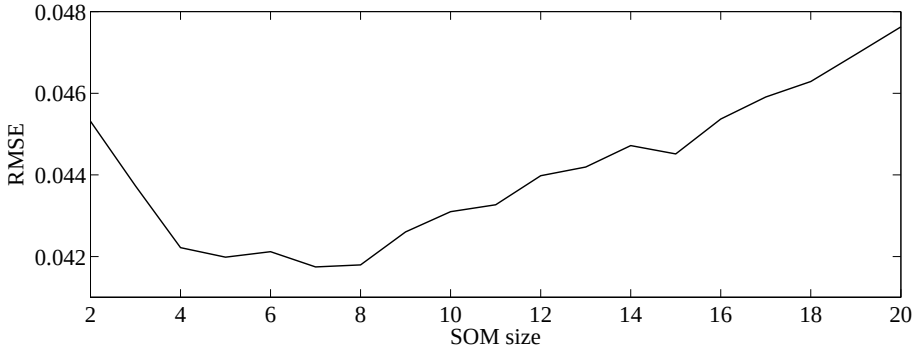


Figure 12. *Validation Errors with Respect to the size of the SOM grid.*

From Figure 12 we can see that the optimal grid size for the SOM is 7×7 giving a RMSE of .0417. This time the optimal number of SOM nodes 42 is clearly smaller than the number of samples 120. Also, the validation error curve is less smooth than with the other datasets, which suggests more difficult dataset to fill.

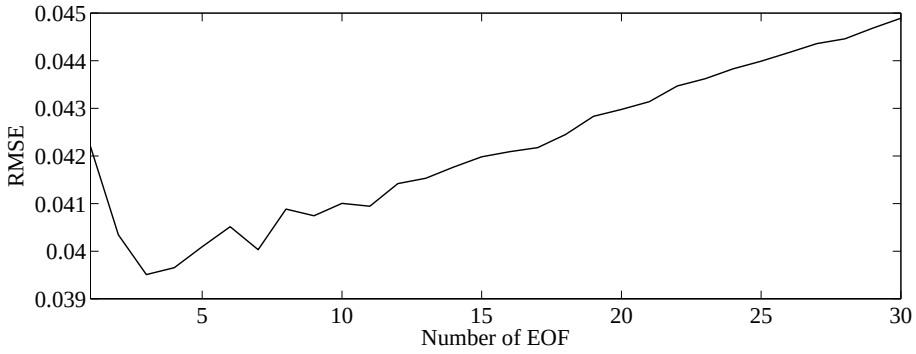


Figure 13. *Validation Errors with Respect to the number of EOF with column mean initialization.*

From Figure 13, we can see that the optimal number of EOF is 3 giving a RMSE of .0395. The number of EOF is very small compared to the number of available EOF 120. This suggests very strong noise influence in the data and is supporting the observation of more difficult dataset to fill.

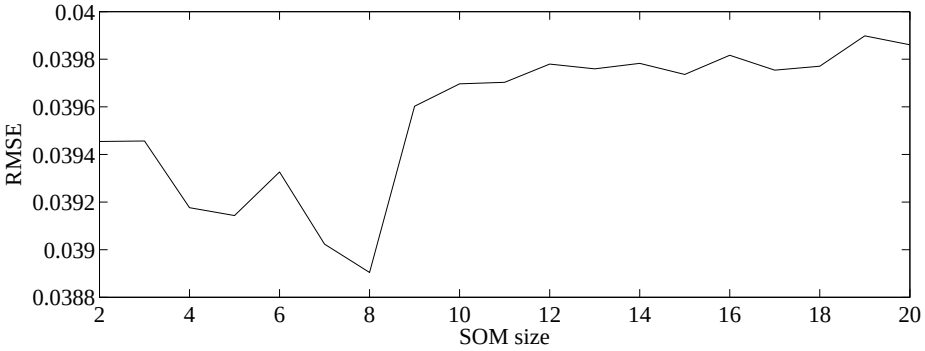


Figure 14. *Validation Errors with Respect to the size of the SOM grid.*

From Figure 14, the optimal SOM size is found to be 8×8. It is roughly the same than using SOM alone, but because of the EOF methodology performed after SOM initialization, the obtained error is lower, .0389.

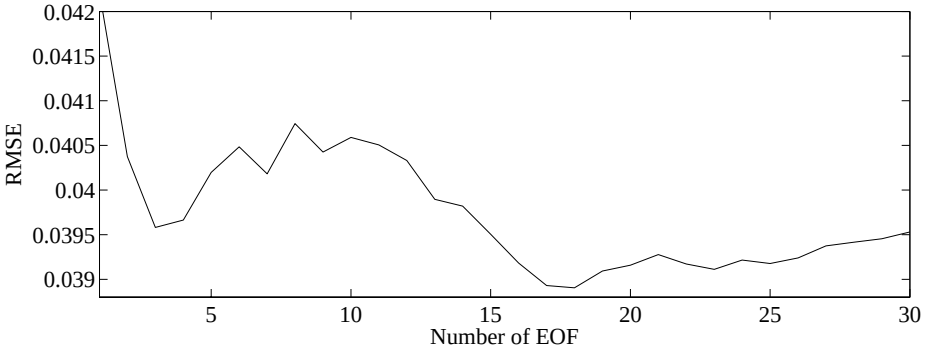


Figure 15. *Validation Errors with Respect to the number of EOF with SOM initialization using a grid size 8×8.*

From Figure 15, the optimal EOF is found to be 18, which is 6 times larger when using EOF alone. This clearly demonstrates the necessity of a good initialization for the EOF, and SOM can provide that.

Also, it can be noted that the number of EOF selected for the EOF alone is a clear local minimum even when using SOM initialization. It suggests that not all 18 EOF are necessary, even though the global optimum is obtained, but instead some of the EOF between 3 and 18 could be pruned out.

Table 6. *Validation and test RMS errors for all the methods using a publicly available financial dataset.*

10^{-2}	Validation Error	Test Error
ECM	4.18	4.34
SOM	4.17	3.92
EOF	3.95	3.88
SOM + EOF	3.89	3.60

Comparing the validation and test errors in the Table 6, the SOM+EOF clearly outperforms the other methodologies. Since this dataset is the hardest one, all the errors are also larger by one order of magnitude compared to the American and European fund return datasets.

6. Conclusion

In this paper, we have compared various methods for finding missing values in temporal databases. The methods are Self-Organizing Maps (SOM), Empirical Orthogonal Function (EOF), the combination of the two, the SOM+EOF method and the Expectation Conditional Maximisation Method (ECM) used as benchmark.

The advantages of the SOM include the ability to perform nonlinear projection of high-dimensional data to lower dimension with interpolation between discrete data points.

For the EOF, the advantages include high-dimensional linear projection of high-dimensional data and the speed and the simplicity of the method.

The combination of the two methods include the advantages of both individual methods, leading to a new accurate approximation methodology for finding the missing values. The several test performed shows the accuracy of the new methodology for completing missing values.

It has also been shown experimentally that the optimal number of code vectors used in the SOM has to be larger than the number of observations. It is necessary in order to take the advantage of the self-organizing property of the SOM and the interpolation ability for finding the missing data.

Furthermore, the amount of missing values is neither restricting the usage of the method nor seriously decreasing the performance.

For further works, the modifications and performance upgrades for the global methodology are fine-tuned for different types of datasets. The methodology will then be applied to datasets from other field of science, for example climatology and process data.

Bibliography

[BOY 94] BOYD J., KENNELLY E., PISTEK P., "Estimation of eof expansion coefficients from incomplete data", *Deep-sea research*, vol. 41, num. 10, 1994, p. 1479-1488.

[DEM 77] DEMPSTER A., LAIRD N., RUBIN D., "Maximum likelihood from incomplete data *via* EM algorithm", *Journal of the Royal Statistical Society*, vol. 39, 1977, p. 1-38.

[FES 02] FESSANT F., MIDENET S., "Self-organizing map for data imputation and correction in survey", *Neural Computation and Application*, vol. 10, 2002, p. 300-310.

[KOH 95] KOHONEN T., *Self-Organizing Maps*, Springer-Verlag, Berlin, 1995.

[LEN 03] LENDASSE A., WERTZ V., VERLEYSEN M., "Model selection with cross-validations and bootstraps - application to time series prediction with rbfn models", *LNCS*, vol. 2714, 2003, p. 573-580.

[MEN 93] MENG X., RUBIN D., "Maximum likelihood estimation *via* the ECM algorithm", *Biometrika*, vol. 80, 1993, p. 267-278.

[PRE 88] PREISENDORFER R., *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, 1988.

[SAM 92] SAMAD T., HARP S., "Self organization with partial data", *Network*, vol. 3, 1992, p. 205-212.

[URL 01] SOM TOOLBOX: HTTP://WWW.CIS.HUT.FI/PROJECTS/SOMTOOLBOX/

[URL 02] SOM+EOF TOOLBOX: HTTP://WWW.CIS.HUT.FI/PROJECTS/TSP/INDEX.PHP?PAGE=RESEARCH&SUBPAGE=DOWNLOADS

[WAN 03] WANG S., "Application of self-organising maps for data mining with incomplete data set", *Neural Computing and Applications*, vol. 12, num. 1, 2003, p. 42-48.